



# Synthèse de parole expressive au delà du niveau de la phrase : le cas du conte pour enfant : conception et analyse de corpus de contes pour la synthèse de parole expressive

David Doukhan

## ► To cite this version:

David Doukhan. Synthèse de parole expressive au delà du niveau de la phrase : le cas du conte pour enfant : conception et analyse de corpus de contes pour la synthèse de parole expressive. Informatique et langage [cs.CL]. Université Paris Sud - Paris XI, 2013. Français. NNT : 2013PA112165 . tel-01161731

**HAL Id: tel-01161731**

**<https://theses.hal.science/tel-01161731>**

Submitted on 9 Jun 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS SUD

Spécialité : Informatique

Présentée par

**David Doukhan**

Pour obtenir le titre de  
DOCTEUR de L'UNIVERSITÉ PARIS 11

Sujet :

---

***Synthèse de parole expressive au delà du niveau de la  
phrase***

***Le cas du conte pour enfant.***

Conception et analyse de corpus de contes pour la synthèse de parole expressive

---

préparée

au LABORATOIRE D'INFORMATIQUE POUR LA MÉCANIQUE ET LES SCIENCES DE L'INGÉNIEUR  
(CNRS UPR 3251)

Soutenue le 20 septembre 2013 devant le jury composé de

Christophe D'ALESSANDRO	Directeur	LIMSI-CNRS
Olivier BOËFFARD	Rapporteur	IRISA/ENSSAT
Thierry DUTOIT	Rapporteur	FPMs
Hansjörg MIXDORFF	Examineur	BHT Berlin
Albert RILLIARD	Co-encadrant	LIMSI-CNRS
Sophie ROSSET	Examinatrice	LIMSI-CNRS



---

## Remerciements

La rédaction de ce manuscrit est le résultat de quatre années de thèse, que j’ai eu le plaisir de préparer au sein de l’équipe Audio-Acoustique du LIMSI. La description des phénomènes prosodiques qui fut l’une des thématiques centrales de ce manuscrit, est souvent déroutante de par sa complexité. L’achèvement de ce manuscrit a été rendu possible par une combinaison de rencontres, de discussions, de débats, de retours d’expérience, et de remises en question. Je tiens donc à remercier dans cette partie tous ceux qui ont pu y contribuer.

Mes premiers remerciements vont à mes deux encadrants : Christophe d’Alessandro et Albert Rilliard. Tous deux on été très impliqués et disponibles pendant l’ensemble de la durée de mon doctorat. Ils m’ont fait bénéficier de leur vaste expérience et m’ont apporté un soutien sans failles tout au long de mon cursus. J’ai pu ainsi bénéficier d’un encadrement scientifique rigoureux, dans un cadre de travail particulièrement agréable. Je tiens à remercier Christophe de sa patience, ainsi que de m’avoir poussé à finaliser la rédaction de mes articles et du présent manuscrit. Je ne saurais que trop remercier Albert pour l’aide considérable qu’il m’a apporté tout au long du doctorat, qu’il s’agisse de relectures, de corrections, d’analyse de données ou encore d’initiation aux démarches descriptives.

Je tiens à remercier les rapporteurs de cette thèse : Thierry Dutoit et Olivier Boëffard - ainsi que les examinateurs Hansjörg Mixdorff et Sophie Rosset - d’avoir accepté de faire partie du jury et de m’avoir fait remonter un grand nombre de précieux commentaires.

Plusieurs personnes ont contribué à m’orienter sur des thématiques qui n’entraient pas dans les domaines de compétences privilégiés de mes encadrants. Je tiens donc à remercier en premier lieu Sophie Rosset pour son aide qui fut décisive sur l’ensemble des travaux ayant trait à la linguistique, au traitement automatique des langages et à l’analyse des accords inter-annotateurs ; Martine Adda-Decker pour son aide en phonétique et en alignement semi-automatique de la parole ; Thomas Lavergne pour ses explications sur les fondements théoriques relatifs aux CRF et sur l’utilisation de l’outil Wapiti ; Laurent Pointal pour m’avoir orienté vers des outils python correspondant à mes besoins ; Abderahman Azhar pour m’avoir transmis les bases permettant d’utiliser un cluster de calcul ; Nicolas Foucault pour les échanges d’expérience concernant les stratégies de segmentation de texte ; Pépito Matéo, pour avoir accepté ma présence au sein de son cours “Leçon d’oralité” et avoir contribué à ma sensibilisation au métier de conteur.

Je tiens également à remercier les personnes qui ont rendu le chemin du BAC au doctorat possible : Aymeric Zils pour m’avoir permis de développer des compétences en modélisation prosodique qui ont fortement joué sur mon recrutement au sein de l’équipe Audio-Acoustique ; Anne Sédès et Olivier Warusfel pour m’avoir encouragé et guidé lors de la rédaction de mon premier article ; Christian Dujardin et Akli Adjaoute pour m’avoir accordé leur pleine confiance lors de la mise en place de modules d’enseignements à l’EPITA ; Olivier Rodot et Christophe Boullay pour m’avoir fortement encouragé à faire de la recherche et de l’enseignement, dans un contexte où je n’avais jamais été sensibilisé à ce type de carrière ; Reda Dehak pour m’avoir initié au traitement



---

du signal audio durant son temps libre ; Yves-Jean Daniel, Alain Tissier, Nicolas Clauvelin et Benoit Monin d'avoir réussi à me réconcilier avec les mathématiques.

Je tiens à remercier mes collègues de l'équipe AA pour les bons moments passés en leur compagnie : Tifanie de m'avoir autorisé à dévaliser quotidiennement son stock de sucreries, et de m'avoir fait intégrer l'AFREUBO ; Lionel pour avoir été mon frère de master, puis de stage et enfin de thèse, et m'avoir présenté l'ensemble de ses frères musiciens ; Marc R. pour son positivisme à toute épreuve, de m'avoir engagé en coach endurance et fait découvrir des compétences professionnelles inexploitées jusqu'ici ; Marc E. pour son soutien lors de la dernière ligne droite de la présentation orale du manuscrit, les corrections typographiques de dernière minute sur mes transparents et l'aide technique lors de la soutenance ; Guillaume et Johan pour m'avoir fréquemment sollicité pour des formations G-Switch et autres recadrages ; Sylvain pour le coaching en début de thèse ; Nicolas A., parce que ça fait plaisir de tomber sur des collègues calés sur l'historique du punk-rock ; David PQ pour avoir vaillamment essayé d'utiliser mon external PureData ; Paul pour sa patate et Nathalie pour ses retours d'expérience pédagogique.

Toute cette période n'aurait pas été aussi enrichissante sans la présence de mes chers colocataires : Éric, Gérard, Ignace, Nicole. Sans oublier tous les membres des groupes où j'ai pu officier pendant la durée de la thèse : les babouches, l'afreubo, bombo, la grâce de l'hippopotame, les mouettes & charbons et la Fabrika Skank.

Pour finir, je tiens à remercier mes amis de m'avoir supporté pendant ces quatre années, mes parents qui ont fait tout leur possible pour que je puisse évoluer dans la voie que j'ai choisie, mon frère et ma sœur, ma tante qui a acceptée d'être de corvée orthographique, mon oncle et mes grands-parents pour leur soutien quotidien.

---

## Résumé

L’objectif de la thèse est de proposer des méthodes permettant d’améliorer l’expressivité des systèmes de synthèse de la parole. Une des propositions centrales de ce travail est de définir, utiliser et mesurer l’impact de structures linguistiques opérant *au delà du niveau de la phrase*, par opposition aux approches opérant sur des phrases isolées de leur contexte.

Le cadre de l’étude est restreint au cas de la lecture de contes pour enfants. Les contes ont la particularité d’avoir été l’objet d’un certain nombre d’études visant à en dégager une structure narrative et de faire intervenir un certain nombre de stéréotypes de personnages (héros, méchant, fée) dont le discours est souvent rapporté. Ces caractéristiques particulières sont exploitées pour modéliser les propriétés prosodiques des contes *au delà du niveau de la phrase*. La transmission orale des contes a souvent été associée à une pratique musicale (chants, instruments) et leur lecture reste associée à des propriétés mélodiques très riches, dont la reproduction reste un défi pour les synthétiseurs de parole modernes.

Pour répondre à ces problématiques, un premier corpus de contes écrits est collecté et annoté avec des informations relatives à la structure narrative des contes, l’identification et l’attribution des citations directes, le référencement des mentions des personnages ainsi que des entités nommées et des énumérations étendues. Le corpus analysé est décrit en terme de couverture et d’accord inter-annotateurs. Il est utilisé pour modéliser des systèmes de segmentation des contes en épisodes, de détection des citations directes, des actes de dialogue et des modes de communication. Un deuxième corpus de contes lus par un locuteur professionnel est présenté. La parole est alignée avec les transcriptions lexicale et phonétique, les annotations du corpus texte et des méta-informations décrivant les caractéristiques des personnages intervenant dans le conte. Les relations entre les annotations linguistiques et les propriétés prosodiques observées dans le corpus de parole sont décrites et modélisées. Finalement, un prototype de contrôle des paramètres expressifs du synthétiseur par sélection d’unités Acapela est réalisé. Le prototype génère des instructions prosodiques opérant *au delà du niveau de la phrase*, notamment en utilisant les informations liées à la structure du conte et à la distinction entre discours direct et discours rapporté. La validation du prototype de contrôle est réalisée dans le cadre d’une expérience perceptive, qui montre une amélioration significative de la qualité de la synthèse.

## Mots clés

Synthèse de parole, corpus de texte, corpus de parole, expressivité, prosodie, accord inter-annotateur, conte pour enfant, structure narrative, citation directe, CRF

---

## Abstract

The aim of this thesis is to propose ways to improve the expressiveness of speech synthesis systems. One of the central propositions of this work is to define, use and measure the impact of linguistic structures operating beyond the sentence level, as opposed to approaches operating on sentences out of their context. The scope of the study is restricted to the case of storytelling for children. The stories have the distinction of having been the subject of a number of studies in order to highlight a narrative structure and involve a number of stereotypical characters (hero, villain, fairy) whose speech is often reported. These special features are used to model the prosodic properties tales beyond the sentence level. The oral transmission of tales was often associated with musical practice (vocals, instruments) and their reading is associated with rich melodic properties including reproduction remains a challenge for modern speech synthesizers. To address these issues, a first corpus of written tales is collected and annotated with information about the narrative structure of stories, identification and allocation of direct quotations, referencing references to characters as well as named entities and enumerations areas. The corpus analyzed is described in terms of coverage and inter-annotator agreement. It is used to model systems segmentation tales episode, detection of direct quotes, dialogue acts and modes of communication. A second corpus of stories read by a professional speaker is presented. The word is aligned with the lexical and phonetic transcriptions, annotations of the corpus text and meta-information describing the characteristics of the characters involved in the story. The relationship between linguistic annotations and prosodic properties observed in the speech corpus are described and modeled. Finally, a prototype control expressive synthesizer parameters by Acapela unit selection is made. The prototype generates prosodic operating instructions *beyond the sentence level*, including using the information related to the structure of the story and the distinction between direct speech and reported speech. Prototype validation control is performed through a perceptual experience, which shows a significant improvement in the quality of the synthesis.

## Keywords

Text To Speech, text corpus , speech corpus, expressivity, prosody, inter annotator agreement, children tale, narrative structure, direct quotation , CRF

---

*Sa parole (du conteur) est éphémère : belle et juste dans l'instant, elle s'appauvrit bien souvent lorsqu'on la retire de son contexte, dans l'espoir fébrile de l'exploiter. Ou de l'offrir comme les fleurs. Elle est reliée au moment exact qui l'a fait naître, elle respire avec le partenaire qui la désirait.*

Michel Hindenoch ([Touati, 2000](#), page 24)

---

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	L’art de raconter des histoires . . . . .	1
1.2	Contexte : Le projet GV-LEx . . . . .	3
1.3	Objectifs . . . . .	4
1.4	Démarche . . . . .	5
1.5	Organisation du manuscrit . . . . .	6
<b>2</b>	<b>Conception et Analyse d’un corpus textuel de contes</b>	<b>9</b>
2.1	Résumé du chapitre . . . . .	10
2.2	État de l’art : les corpus de contes annotés . . . . .	10
2.3	Constitution du corpus GV-LEx . . . . .	14
2.3.1	Collecte de contes . . . . .	14
2.3.2	Normalisation des textes . . . . .	15
2.3.3	Protocole d’annotation . . . . .	16
2.4	Schéma d’annotation . . . . .	18
2.4.1	Énumérations étendues . . . . .	19
2.4.2	Entités nommées étendues . . . . .	19
2.4.3	Tours de parole . . . . .	20
2.4.4	Structure narrative . . . . .	20
2.5	Couverture et Répartition des Annotations . . . . .	22
2.5.1	Unités textuelles . . . . .	22
2.5.2	Annotations anaphoriques et chaînes de coréférence . . . . .	24

2.6	Accord inter-annotateurs . . . . .	26
2.6.1	Segmentation en épisodes . . . . .	26
2.6.2	Identification des éléments structurels et lexicaux . . . . .	29
2.6.3	Référencement des personnages . . . . .	30
<b>3</b>	<b>Module d'extraction d'informations linguistique</b>	<b>35</b>
3.1	Résumé du chapitre . . . . .	36
3.2	Annotation des unités lexicales . . . . .	36
3.2.1	Signes de Ponctuation . . . . .	36
3.2.2	Parties du Discours . . . . .	37
3.2.3	Lemmatisation et Racinisation . . . . .	38
3.2.4	Synsets et Hyperonymes . . . . .	39
3.2.5	Structure Syntaxique . . . . .	39
3.2.6	Transcription phonétique . . . . .	41
3.3	Module de détection automatique de tours de parole . . . . .	41
3.3.1	Modèle de détection basé sur des automates simples . . . . .	42
3.3.2	Modèles basés sur des Champs Conditionnels Aléatoires Linéaires . . . . .	43
3.3.3	Évaluation . . . . .	47
3.3.4	Conclusion Partielle . . . . .	48
3.4	Modèles de segmentation en épisodes . . . . .	50
3.4.1	Modèles de segmentation en épisodes basés sur des automates . . . . .	51
3.4.2	Modèles des segmentation statistiques . . . . .	52
3.4.3	Conclusion partielle . . . . .	55
3.5	Module d'extraction des actes de dialogue et de modes de communication . . . . .	56
3.5.1	Modes de communication . . . . .	56
3.5.2	Actes de dialogue . . . . .	57
3.5.3	Conclusion partielle . . . . .	57

<b>4</b>	<b>Conception et Analyse d'un corpus oral de contes</b>	<b>59</b>
4.1	Corpus oraux de contes annotés . . . . .	60
4.2	Enregistrement d'un corpus de contes . . . . .	61
4.3	Alignement lexical et phonétique de la parole . . . . .	62
4.3.1	Transcription lexicale de la parole . . . . .	62
4.3.2	Procédure d'alignement . . . . .	64
4.4	Annotations . . . . .	66
4.5	Stylisation Prosodique . . . . .	67
4.5.1	Estimation de descripteurs prosodiques bruts . . . . .	70
4.5.2	Stylisation basée sur un modèle de perception tonale . . . . .	71
4.5.3	Enregistrement des données stylisées dans des feuilles de calcul . . . . .	71
4.5.4	Représentation graphique du corpus GV-LEx . . . . .	72
4.5.5	Stylisation basée sur le Prosogram . . . . .	73
<b>5</b>	<b>Analyse Prosodique</b>	<b>75</b>
5.1	Introduction . . . . .	75
5.2	Descripteurs prosodiques . . . . .	76
5.3	Comparaison des propriétés prosodiques globales du corpus de contes à d'autres corpus de parole . . . . .	77
5.4	Propriétés prosodiques liées aux étiquettes d'épisodes . . . . .	78
5.5	Caractéristiques prosodiques liées à l'incarnation des personnages . . . . .	80
5.5.1	Impact de l'âge et du genre sur les distributions de fréquence fondamentale et d'intensité . . . . .	80
5.5.2	Regroupement des personnages basé sur leur profil prosodique . . . . .	82
5.6	Analyse et modélisation de la durée des pauses . . . . .	89
5.6.1	Motivations . . . . .	89
5.6.2	Méthodologie . . . . .	89
5.6.3	Distributions des pauses . . . . .	90
5.6.4	Modélisation non paramétrique de la durée des pauses en fonction de la taille du contexte . . . . .	91



5.6.5	Modèles paramétriques . . . . .	92
5.6.6	Impact des tours de parole sur la durée des pauses . . . . .	96
5.7	Conclusion partielle . . . . .	98
<b>6</b>	<b>Un prototype de synthèse de parole expressive</b>	<b>101</b>
6.1	Résumé du chapitre . . . . .	102
6.2	Architecture du système de synthèse de parole expressive . . . . .	102
6.3	Caractéristiques du synthétiseur de parole utilisé . . . . .	103
6.3.1	Instructions prises en charge . . . . .	103
6.3.2	Voix de synthèse . . . . .	105
6.3.3	Outil d'édition manuelle des paramètres expressifs utilisés pour la synthèse	106
6.4	Module d'annotation prosodique automatique . . . . .	107
6.5	Méthodologie d'évaluation du prototype de synthèse . . . . .	110
6.6	Analyse des résultats du test perceptif . . . . .	113
6.6.1	Évaluations réalisées par les enfants . . . . .	113
6.6.2	Évaluations réalisées par les adultes . . . . .	115
<b>7</b>	<b>Conclusion</b>	<b>121</b>
7.1	Contributions de la thèse . . . . .	122
7.1.1	Constitution, évaluation et diffusion de corpus . . . . .	122
7.1.2	Extraction automatique d'informations dans des textes fictifs . . . . .	123
7.1.3	Description des propriétés prosodiques observées dans les contes lus . . . . .	123
7.1.4	Prototype de synthèse de parole expressive . . . . .	124
7.2	Perspectives de recherche . . . . .	124
7.2.1	Extension du schéma d'annotation proposé . . . . .	124
7.2.2	Mise en application de l'ensemble des règles prosodiques formulées . . . . .	125
7.2.3	Raffinement des modèles de prédiction de la durée des pauses . . . . .	126
7.2.4	Vers une reformulation du texte . . . . .	126
7.3	Publications liées à la thèse . . . . .	128
7.3.1	Article de revue à comité de lecture . . . . .	128

7.3.2	Conférences avec actes . . . . .	128
7.3.3	Communication sans actes . . . . .	128
<b>Bibliographie</b>		<b>128</b>
<b>A Document Type Definition (DTD) utilisée pour l’annotation du corpus texte</b>		<b>141</b>
<b>B Exemple de conte annoté : Le petit chaperon rouge</b>		<b>145</b>
<b>C Modélisation CRF des tours de parole</b>		<b>165</b>
C.1	Données d’entrée . . . . .	165
C.2	Patron d’extraction de fonction features . . . . .	177
<b>D Propriétés prosodiques associées aux personnages</b>		<b>183</b>
<b>E Textes sélectionnés pour l’évaluation du système de synthèse</b>		<b>187</b>
E.1	Erik , le paysan rusé . . . . .	187
E.2	Crétin de paon . . . . .	187
E.3	Naëlle va chercher du bois dans la forêt . . . . .	188
E.4	Le roi Glagla . . . . .	188
E.5	L’histoire des trois sourds . . . . .	189
E.6	Maxence et le monstre sous le lit . . . . .	189
<b>F Stylistique du conte “Le petit chaperon rouge” contenu dans le corpus de parole GV-LEx</b>		<b>191</b>



# Table des figures

1.1	Le robot humanoïde Nao . . . . .	3
1.2	L’avatar Greta développé par Télécom ParisTech . . . . .	4
2.1	Capture d’écran du logiciel d’annotation linguistique conçu par l’entreprise Syllabs pour les besoins spécifiques du projet GV-LEx. . . . .	17
3.1	Arbre syntaxique abstrait correspondant à la phrase “le chat mange la souris” . . .	40
3.2	Arbre syntaxique abstrait équivalent à la figure 3.1 représenté à l’aide d’un attribut multidimensionnel par mot . . . . .	40
3.3	Variation des performances obtenues pour la tâche de segmentation en tours de parole par les modélisations basées sur des CRF, pour des tailles contextuelles d’unigrammes et de bigrammes fixées . . . . .	49
4.1	Transcription de l’enregistrement d’un conte à l’aide du logiciel Transcriber. . . . .	63
4.2	Visualisation et édition des frontières lexicales et phonétiques depuis Praat . . . . .	65
4.3	Exemple réduit d’annotations alignées avec le signal de parole et visualisées à l’aide du logiciel Praat. . . . .	69
4.4	Représentation graphique de la prosodie du corpus GV-LEx, stylisée selon un modèle de perception tonale. . . . .	73
4.5	Stylisation obtenue avec le Prosogram . . . . .	74
5.1	Distribution de F0 et d’intensité observées pour les différents rôles narratifs du conteur : narrateur (N0) et différentes catégories de personnages regroupées par âge (enfants, adolescents, adultes, âgé) et par genre (Féminin/Masculin) . . . . .	81
5.2	Dendrogramme montrant le regroupement hiérarchique des personnages féminins, obtenu à partir de l’ACP menée sur les données issues de la stylisation en Prosogram	85

5.3	Dispersion des personnages féminins sur les deux premières composantes de l'ACP montrant leurs différences en fonction des paramètres prosodiques extraits de la stylisation par Prosogram . . . . .	85
5.4	Dendrogramme montrant le regroupement hiérarchique des personnages masculins obtenu à partir de l'ACP menée sur les données issues de la stylisation en Prosogram . . . . .	88
5.5	Dispersion des personnages masculins sur les deux premières composantes de l'ACP montrant leurs différences en fonction des paramètres prosodiques extraits de la stylisation par Prosogram . . . . .	88
5.6	Durées des différents types de pauses : à l'intérieur des pseudo-syntagmes, aux frontières des pseudo-syntagmes, des phrases et des paragraphes. . . . .	91
5.7	Durée des pauses observées aux frontières des phrases ( $\log_{10}(ms)$ ), obtenue à l'aide d'un modèle de régression à base de <i>Kernel</i> , utilisant la taille des phrases précédentes et suivantes. . . . .	93
5.8	Durée des pauses observées aux frontières des pseudo-syntagmes ( $\log_{10}(ms)$ ), obtenue à l'aide d'un modèle de régression à base de <i>Kernel</i> , utilisant la taille des pseudo-syntagmes précédents et suivants. . . . .	93
6.1	Architecture du système de synthèse de parole expressive . . . . .	104
6.2	Interface d'édition des propriétés expressives globales de la synthèse depuis Virtual Story Teller, logiciel d'édition manuelle d'instructions prosodiques développé par Acapela Group . . . . .	107
6.3	Interface d'édition des propriétés expressives locales de la synthèse depuis Virtual Story Teller, logiciel d'édition manuelle d'instructions prosodiques développé par Acapela Group . . . . .	108
6.4	Instructions fournies aux sujets lors de l'évaluation du système de synthèse expressive . . . . .	111
6.5	Interface d'évaluation du système de synthèse expressive . . . . .	112
6.6	Dispersion des notes données à chacun des 4 systèmes par les sujets de moins de 10 ans. . . . .	113
6.7	Dispersion des notes données par les enfants à la synthèse neutre (SN), à la synthèse automatique (SA), à la synthèse manuelle (SM) et au locuteur professionnel (LP), pour chaque conte . . . . .	114
6.8	Comparaison de la dispersion des notes données par les enfants et les adultes . . . . .	115
6.9	Évaluation de la qualité des systèmes de synthèse . . . . .	116
6.10	Évaluation de la qualité des systèmes de synthèse par conte . . . . .	117

# Liste des tableaux

2.1	Caractéristiques de corpus de contes annotés décrits dans la littérature . . . . .	11
2.2	Nombre moyen, minimum et maximum d'éléments observés sur les 89 contes suite au processus de tokenisation . . . . .	15
2.3	Marqueurs XML utilisés dans les corpus GV-LEx . . . . .	18
2.4	Distribution des annotations dans les 89 contes du corpus texte . . . . .	23
2.5	Description des chaines de coréférence associées aux tours de parole (TP), aux entités personne parlante (EPP) et aux entités personne non parlante (EPNP) . .	25
2.6	Matrice de confusion entre A1 et A2 pour la segmentation en épisodes. . . . .	27
2.7	Table d'interprétation du Kappa de Cohen . . . . .	27
2.8	Accord sur la localisation des frontières d'épisodes entre A1 et A2, estimé à l'aide des métriques Kappa, F-Measure, WindowDiff, Weighted WindowDiff et Pk . . . .	28
2.9	Estimation de l'accord inter-annotateurs pour les unités lexicales et les tours de parole. . . . .	31
2.10	Accord inter-annotateurs estimé pour les annotations anaphoriques . . . . .	33
3.1	Accord observé pour l'étiquetage des parties du discours entre un système à base de HMM et un système basé sur TreeTagger . . . . .	38
3.2	Comparaison des racines obtenues avec NLTK et des lemmes obtenus avec TreeTagger	39
3.3	Résultats obtenus par les automates et les modélisations à base de CRF pour la tâche de segmentation en tours de parole. . . . .	49
3.4	Évaluation des modèles de segmentation en épisodes basés sur des automates . . .	51
3.5	Évaluation de modèles de segmentation en épisodes basés sur TextTiling et optimisant la F-Mesure . . . . .	54
3.6	Évaluation de modèles de segmentation en épisodes basés sur TextTiling et optimisant WindowDiff . . . . .	54

3.7	Évaluation de modèles de segmentation en épisodes basés sur TextTiling et optimisant Pk . . . . .	54
4.1	Corpus annotés de parole contenant des contes . . . . .	61
4.2	Caractéristiques des transcriptions des douze contes contenus dans le corpus de parole . . . . .	64
4.3	Différences observées entre les textes des contes fournis au locuteur (9664 mots) et les transcriptions des enregistrements du corpus de parole . . . . .	64
4.4	Description des données obtenues suite à l'alignement lexical et phonétique du corpus de parole (durée totale : 3593 secondes) . . . . .	66
4.5	Méta-informations utilisées pour l'enrichissement de la caractérisation des personnages du corpus de parole . . . . .	67
4.6	Couverture des annotations linguistiques pour les douze contes du corpus de parole . . . . .	68
5.1	Comparaison des propriétés prosodiques du corpus GV-LEx et du corpus décrit par Patel et al. (2006) . . . . .	77
5.2	Propriétés prosodiques des corpus GV-LEx et GV-LEx-Narr et des styles correspondant aux actualités radiophoniques (AR), à des discours politiques (DP) et à des conversations (CONV) (Roekhaut et al., 2010) . . . . .	78
5.3	Comparaison des propriétés prosodiques associées aux étiquettes d'épisodes annotés dans le corpus GV-LEx . . . . .	79
5.4	Données de l'Analyse en Composantes Principale menée sur le groupe de personnages féminins . . . . .	84
5.5	Données de l'Analyse en Composantes Principale menée sur le groupe de personnages masculins . . . . .	87
5.6	Distributions des pauses observées dans le corpus GV-LEx . . . . .	90
5.7	Modèles paramétriques utilisés pour la prédiction de la durée des pauses en fonction de la taille du contexte . . . . .	94
5.8	Erreur quadratique moyenne obtenue par les modèles paramétriques de prédiction de la durée d'une pause en fonction de la taille du contexte. . . . .	95
5.9	Influence des tours de parole sur la durée des pauses inter-phrase . . . . .	97
5.10	Influence des tours de parole sur la durée des pauses inter-pseudo-syntagme . . . . .	98
6.1	Règles de mise en relation des informations linguistiques avec le choix de la base d'unités de synthèse . . . . .	109
6.2	Résultats de l'ANOVA appliquée aux réponses des sujets adultes . . . . .	115

6.3	Résultats du test HSD de Tukey mesurant la signification statistique des différences entre les différents niveaux des facteurs <i>système</i> (pour les systèmes automatiques) et <i>conte</i> . . . . .	118
D.1	Propriétés prosodiques associées aux personnages . . . . .	184
D.1	Propriétés prosodiques associées aux personnages . . . . .	185





# Chapitre 1

## Introduction

### 1.1 L’art de raconter des histoires

L’art de raconter des histoires est l’une des plus vieilles formes artistiques, dont découlent le théâtre et la littérature (Henri Gougaud dans [Touati, 2000](#)). Les contes sont intimement liés à la *tradition orale* et ont eu pour fonctions l’éducation des enfants, la transmission de valeurs morales, la prévention des dangers et la construction d’histoires collectives. À l’aide d’*images*, ils permettent de répondre aux questions angoissantes que peuvent se poser les enfants et jouent un rôle central dans leur développement ([Bettleheim, 1976](#)).

Le métier de conteur a pris différentes formes dans les sociétés traditionnelles, qu’il s’agisse des aèdes de la Grèce antique, des bardes de l’antiquité celte, des troubadours trouvères et ménestrels de l’Europe moyenâgeuse, des griots africains ou des geishas au Japon. En fonction du contexte culturel considéré, les conteurs pouvaient être amenés à transmettre les histoires sous forme de chants ou en s’accompagnant d’un instrument de musique. Au quotidien, la fonction de conteur reste bien souvent assurée par les parents ou la famille proche.

On recense en France trois cents conteurs professionnels ayant un statut d’intermittent du spectacle et près de quatre mille personnes ayant une activité de conteur amateur ([Touati, 2000](#)). Chaque conteur se caractérise par son répertoire particulier. Les artistes-conteurs définissent leur activité comme l’art de raconter oralement une histoire sans support autre que sa mémoire, son imaginaire et ses talents d’improvisation. Les conteurs sont dans un rapport de proximité avec leur auditoire et interviennent au sein d’écoles, bibliothèques, théâtres, cafés et festivals.

Mis à part les choix lexicaux, la transmission d’images, d’émotions et d’intuitions se fait principalement via un *processus de communication non verbale*, qui peut se décrire par des intonations, des postures et des mimiques. Le contrôle de ces processus de communication est difficile et nécessite

assez souvent que le conteur fasse appel à des procédés d'auto-suggestion comme se persuader temporairement que l'histoire est réelle ou encore qu'il assiste de façon imaginaire à la scène qu'il décrit au cours de sa prestation. Diverses citations, attribuées à des conteurs professionnels, sont présentées ci-dessous, afin d'illustrer les différents procédés utilisés par les conteurs pour donner vie à leurs histoires et ainsi captiver leur auditoire.

*Pour raconter un conte, il faut qu'il me travaille au corps, qu'il ne me lâche plus, que j'en rêve. (...) Je cherche la formulation, les mots, le rythme. C'est toujours très musical. Cela demande beaucoup d'essayages. (...) Quand je raconte une histoire, je la vis, je m'installe au cœur de l'action. Je peux être un simple témoin mais je peux aussi, à un moment donné, incarner un personnage. Je deviens partie prenante de l'histoire si je dis "je", même si ce "je" n'est pas nécessairement un "je" de vérité. Je peux également sortir de l'histoire puis y entrer à nouveau, à la différence d'un acteur qui, lui, joue toujours un personnage.*

Muriel Bloch (Touati, 2000, page 19)

*Je me définis comme un maillon d'une chaîne. J'ai le devoir (...) de nourrir ce maillon qui est l'instant où je raconte une histoire, d'autant de vie que possible. Un conteur nourrit un conte de sa propre vie, de ce qu'il est, de ses expériences, de son talent, de sa capacité de transmettre quelque chose qui est de l'ordre du non-dit et de l'indicible. (...) Si on admet que tous les arts ont une racine commune qui est, comme les chercheurs le disent, la magie primitive, l'art le plus proche de la magie primitive, c'est celui de la parole, celui du prophète, du poète, de l'aède. A la suite de ces prophètes, de ces poètes, de ces aèdes est le conteur.*

Henri Gougoud (Touati, 2000, pages 22-23)

*J'écris des textes et je les raconte dans une adresse directe au public, c'est dans ce sens que je suis conteur, mais en même temps, je me donne la possibilité de jouer en scène les situations, les personnages ou de devenir les morceaux de décor de manière éphémère tout en accordant de l'importance à la forme du langage, au jeu avec les mots. Pour moi, la forme est aussi importante que le fond. Ce qui se raconte me tient à cœur, mais la manière dont c'est raconté et la relation avec le public m'importent autant que l'histoire elle-même. Déjà, lorsqu'un conteur raconte, dans son corps, dans sa voix, dans sa relation à l'auditoire, il manifeste des images par, disons, le théâtre qu'il se fait dans sa tête.*

Pépito Matéo (Touati, 2000, pages 28-29)



FIGURE 1.1 – Le robot humanoïde Nao

## 1.2 Contexte : Le projet GV-LEx

La thèse présentée dans ce manuscrit est réalisée dans le cadre du projet GV-LEx (Geste et Voix pour une Lecture Expressive [Gelin et al., 2010](#)). Le projet fait intervenir trois entreprises et deux laboratoires de recherche : Aldebaran Robotics, Acapela group, Syllabs, Télécom Paris-Tech et le LIMSI-CNRS. Il a pour objectif de doter le robot humanoïde Nao ([Gouaillier et al., 2009](#)) et l’avatar Greta ([Bevacqua et al., 2007](#)) de capacités expressives suffisantes pour lire des histoires face à un public d’enfants.

Comme cela a été mentionné dans la section précédente, le *racontage* est une discipline artistique à part entière. La compréhension, l’imprégnation et l’adhésion à l’histoire dans son ensemble est la condition nécessaire pour en faire une interprétation optimale. Ce projet se restreint à la *lecture expressive* de textes et n’aborde pas un certain nombre de problématiques liées à la profession des conteurs, telles que l’écriture d’histoires originales, la capacité à reformuler des phrases pour pouvoir les prononcer avec un ton optimal, l’improvisation et l’interaction avec le public. Le but du projet consiste donc à créer des systèmes capables d’analyser des textes, pour pouvoir déterminer et retranscrire l’ensemble des intonations, gestes, mimiques appropriés aux différents passages d’un conte.

Dans le cadre de ce projet, Aldebaran assume le rôle de coordination entre les différents partenaires et contribue à l’amélioration des bibliothèques de mouvements pouvant être réalisées par le robot Nao (figure 1.1). La tâche d’Acapela Group consiste à étendre le choix de voix et d’effets pris en charge par leur synthétiseur pour en améliorer les capacités expressives et à réaliser un logiciel permettant d’éditer manuellement les instructions prosodiques prises en charge par le synthétiseur. Le rôle du LIMSI est de réaliser un module permettant d’analyser des textes



FIGURE 1.2 – L’avatar Greta développé par Télécom ParisTech

de contes pour déterminer automatiquement les instructions prosodiques optimales devant être réalisées par le synthétiseur Acapela, ce qui peut être assimilé à un contrôle automatique des paramètres mélodiques et expressifs du synthétiseur. Syllabs a pour mission de réaliser un logiciel d’annotation linguistique spécifique aux besoins du projet et à réaliser l’annotation d’un corpus linguistique selon les consignes d’annotation définies par le LIMSI. Le travail de Télécom ParisTech consiste à utiliser les informations extraites par le module d’analyse du LIMSI pour générer des gestes et des mimiques qui seront exécutés par le robot Nao et l’avatar Greta (figure 1.2).

### 1.3 Objectifs

Le but de la thèse consiste à concevoir un système capable de générer des instructions prosodiques permettant d’améliorer l’expressivité de la parole synthétisée dans le cadre de la lecture de contes. En l’état actuel des connaissances, les systèmes de synthèse de parole se basent principalement sur une analyse syntaxique des phrases pour en déterminer l’intonation. Ces systèmes sont généralement suffisants pour lire des phrases de façon neutre, mais peuvent assez vite provoquer la lassitude de l’auditeur lorsque la taille du texte dépasse quelques paragraphes.

Une des propositions centrales de la thèse est que la prise en compte d’informations linguistiques de haut niveau, notamment des informations *au delà du niveau de la phrase* (Hendricks, 1967; Vandendorpe, 1995; Sluijter and Terken, 2009), peut permettre d’améliorer les capacités expressives des synthétiseurs de parole. Il s’agit donc de :

- Définir des informations linguistiques utiles pour améliorer la synthèse
- Concevoir des systèmes permettant d’extraire automatiquement ces informations linguistiques à partir du texte
- Mettre au point des règles de mise en correspondance des informations linguistiques en instructions prosodiques

## 1.4 Démarche

La démarche proposée pour répondre aux trois problématiques exposées dans la section précédente est fondée sur la constitution et l’analyse de corpus. Un corpus de textes et un corpus de parole sont collectés et annotés pour permettre de modéliser les phénomènes observés au delà du niveau de la phrase.

Le corpus de textes est annoté avec des informations linguistiques susceptibles d’aider à améliorer les systèmes de synthèse de parole et de gestes, telles que la segmentation des contes en épisodes étiquetés, la détection et l’attribution des citations directes, le référencement des personnages, les énumérations et les entités nommées. Il est utilisé à la fois pour analyser la reproductibilité des informations définies et concevoir des systèmes capables d’extraire automatiquement ces informations à partir d’un texte. Le schéma d’annotation défini est suffisamment général pour être utilisé dans d’autres contextes que la synthèse de parole, comme les applications de génération d’histoires (Gervás et al., 2005), la mise au point d’histoires interactives (Grasbon and Braun, 2001) ou encore l’amélioration de la robustesse des systèmes d’extraction d’information, principalement optimisés pour l’analyse des articles de journaux ou des textes médicaux (Goh et al., 2012; Maarouf and Villaneau, 2012).

Le corpus de parole contient des enregistrements de contes lus, annotés avec les informations linguistiques utilisées dans le corpus texte et enrichis avec des annotations permettant de décrire les propriétés prosodiques observées sur le signal. Le corpus est utilisé pour décrire les phénomènes prosodiques observés dans des contes lus, notamment en analysant les interactions de ces phénomènes avec les informations linguistiques de haut niveau associées aux transcriptions.

Les analyses réalisées sur les corpus sont combinées pour proposer un prototype de contrôle automatique des paramètres expressifs du synthétiseur Acapela. Le synthétiseur utilisé est basé sur un moteur de sélection d’unités de taille variable (Bozkurt et al., 2002) et utilise son propre

module d'analyse linguistique pour déterminer l'intonation au niveau de la phrase. Ce type de synthétiseur permet de créer de la parole artificielle de très bonne qualité relativement aux autres méthodes de synthèse existantes, souvent au prix d'un moindre contrôle des paramètres prosodiques et expressifs (Burkhardt and Stegmann, 2009). Le jeu d'instructions permettant de modifier les propriétés prosodiques de la parole synthétisée est relativement restreint, ce qui ne permet de mettre en application qu'une partie des règles prosodiques déterminées lors de l'analyse du corpus de parole. La description prosodique formelle des unités contenues dans les bases n'est pas accessible, et ne permet pas de définir des règles de sélection de bases utilisant uniquement les analyses prosodiques. En conséquences, les règles liées au choix de bases d'unités ont principalement consisté à repérer les motifs lexicaux associés à la description des bases (rires, pleurs, cris, ...).

La validation de l'ensemble de la démarche est réalisée dans le cadre d'une expérience perceptive, consistant à noter des extraits de parole synthétisée obtenus dans différentes conditions.

## 1.5 Organisation du manuscrit

Un certain nombre de thématiques est abordé au long de ce manuscrit : la linguistique de corpus, le traitement automatisé du langage, le traitement du signal, la modélisation prosodique et la synthèse de parole.

Le deuxième chapitre du manuscrit décrit le corpus textuel de contes réalisé dans le cadre du projet. Les spécificités du schéma d'annotation proposé sont détaillées et comparées aux autres corpus de contes annotés mentionnés dans la littérature. La représentativité et la reproductibilité des annotations réalisées sont décrites dans le cadre d'une analyse de couverture et d'accord inter-annotateurs.

Les procédures mises au point pour extraire automatiquement des informations à partir du texte sont décrites dans le troisième chapitre. Elles incluent des procédures d'annotation des unités lexicales, un module de détection des citations directes, un module de segmentation des contes en séquences d'épisodes et un module de détection des actes de dialogue et des modes expressifs.

Le quatrième chapitre décrit le corpus oral de contes. La stratégie d'annotation du corpus est détaillée et comparée aux autres corpus de contes décrits dans la littérature. Des méthodes de stylisation prosodiques mises au point dans le cadre de cette thèse sont détaillées.

Le cinquième chapitre vise à décrire les propriétés prosodiques observées dans le corpus de contes. Les propriétés prosodiques du corpus sont dans un premier temps comparées à des

analyses effectuées dans des conditions similaires sur d'autres styles de lecture afin de déterminer les spécificités de la lecture de contes. Les analyses suivantes décrivent les caractéristiques liées aux différents épisodes et aux caractéristiques des personnages incarnés. Une procédure de modélisation de la durée des pauses en fonction de la taille du contexte est proposée, permettant à la fois d'accroître la variabilité des systèmes de synthèse de parole et de normaliser les durées des pauses dans les corpus oraux.

Le dernier chapitre synthétise l'ensemble des analyses présentées dans ce manuscrit pour proposer un prototype de contrôle automatique des paramètres expressifs de la voix synthétisée. Les règles de mise en relation entre les informations linguistiques de haut niveau et les instructions prosodiques supportées par le synthétiseur Acapela sont détaillées. Le système est évalué à l'aide d'un test perceptif et les réponses des sujets sont analysées.

L'annexe [A](#) détaille la DTD définie pour représenter les informations linguistiques dans le corpus textuel. Un conte entièrement annoté à l'aide de ce schéma est présenté dans l'annexe [B](#). L'annexe [C](#) détaille les modèles conçus pour la détection des citations directes. L'annexe [D](#) fournit les détails des propriétés prosodiques moyennes décrivant les personnages incarnés par le locuteur dans le corpus de parole. L'ensemble des textes utilisés dans le cadre de l'évaluation du prototype de synthèse proposé sont fournis dans l'annexe [E](#). La richesse des phénomènes prosodiques pouvant être observée dans le corpus est illustrée par l'annexe [F](#) et correspond aux représentations graphiques mises au point dans le cadre de cette thèse pour représenter la prosodie.





## Chapitre 2

# Conception et Analyse d'un corpus textuel de contes

### Sommaire

---

<b>2.1</b>	<b>Résumé du chapitre . . . . .</b>	<b>10</b>
<b>2.2</b>	<b>État de l'art : les corpus de contes annotés . . . . .</b>	<b>10</b>
<b>2.3</b>	<b>Constitution du corpus GV-LEx . . . . .</b>	<b>14</b>
2.3.1	Collecte de contes . . . . .	14
2.3.2	Normalisation des textes . . . . .	15
2.3.3	Protocole d'annotation . . . . .	16
<b>2.4</b>	<b>Schéma d'annotation . . . . .</b>	<b>18</b>
2.4.1	Énumérations étendues . . . . .	19
2.4.2	Entités nommées étendues . . . . .	19
2.4.3	Tours de parole . . . . .	20
2.4.4	Structure narrative . . . . .	20
<b>2.5</b>	<b>Couverture et Répartition des Annotations . . . . .</b>	<b>22</b>
2.5.1	Unités textuelles . . . . .	22
2.5.2	Annotations anaphoriques et chaînes de coréférence . . . . .	24
<b>2.6</b>	<b>Accord inter-annotateurs . . . . .</b>	<b>26</b>
2.6.1	Segmentation en épisodes . . . . .	26
2.6.2	Identification des éléments structurels et lexicaux . . . . .	29
2.6.3	Référencement des personnages . . . . .	30

---

## 2.1 Résumé du chapitre

Ce chapitre décrit un corpus de contes s'adressant à des enfants âgés de sept ou huit ans. Le corpus contient quatre-vingt-neuf contes annotés en Français à l'aide de structures linguistiques permettant de décrire le texte au delà du niveau de la phrase : structure narrative, référencement des mentions des personnages, citations directes attribuées. Des structures lexicales sont également utilisées pour identifier les énumérations ainsi que les entités nommées spatiales, temporelles et de personnes. Une analyse des accords inter-annotateurs a été effectuée sur le corpus afin de décrire la reproductibilité du schéma d'annotation proposé. Les accords les plus forts sont observés pour l'identification des citations directes, des entités de personne ainsi que pour les chaînes de références associées aux citations directes et aux entités de personne.

## 2.2 État de l'art : les corpus de contes annotés

Des motivations diverses ont donné lieu à la constitution de corpus de contes annotés.

Parmi les ouvrages les plus influents, on peut mentionner les travaux de [Aarne and Thompson \(1961\)](#), poursuivis par [Uther \(2004\)](#). Ces travaux ont donné lieu à la constitution d'un corpus de plusieurs milliers de contes, organisés de manière thématique. Une structure hiérarchique de *contes types* est définie et utilisée pour classer les contes du corpus. Cette classification a pour but d'isoler les motifs et les variantes culturelles de récits similaires et a permis l'élaboration de plusieurs catalogues de contes.

Dans le cadre de ses travaux sur la morphologie des contes, Vladimir Propp ([Propp, 1928](#)) définit un schéma d'annotation structuraliste restreint aux *contes merveilleux* (classés dans l'index d'Aarne et Thompson sous les numéros 300 à 749) et l'utilise pour annoter une centaine de contes en russe issus des recueils d'Alexandre Afanassiev. Ce formalisme consiste à représenter les contes en fonction de leur structure plutôt qu'en fonction de leurs motifs. Il est pensé pour faciliter l'analyse et la classification des contes merveilleux russes. Les personnages sont assimilés à l'une des sept catégories de personnages abstraits définies : l'agresseur, le donateur, l'auxiliaire, le mandateur, l'objet de la quête, le héros et le faux héros. L'histoire est ensuite retranscrite à l'aide d'une grammaire composée de 31 *fonctions narratives*, regroupées en séquences. Ces fonctions narratives sont créées pour couvrir l'ensemble des actions rencontrées dans les contes traditionnels russes (interdiction, tentative de tromperie, départ du héros...) et ne peuvent être accomplies que par des personnages d'une catégorie donnée. Chaque conte est associé à une séquence de fonctions narratives. La catégorisation des personnages n'est pas fournie explicitement et peut être déduite à partir des séquences narratives. Les portions de texte associées aux fonctions narratives n'y sont pas explicitées.

TABLE 2.1 – Caractéristiques de corpus de contes annotés décrits dans la littérature

Nom	Langue	nb contes	nb phrases	nb mots	type d’annotations
Francisco et al. (2011)	Anglais	18	1389	16 816	états émotionnels
Maarouf and Villaneau (2012)	Français	139		160 000	annotation syntaxique, référencement des personnages
Zhang et al. (2003)	Anglais	133		438 000	citations directes partielles, attribution des citations, entités personne
Mamede and Chaleira (2004)	Portugais	15			citations directes, méta-informations
Goh et al. (2012)	Anglais	8		12 000	Identification des personnages principaux
Malec (2010)	Russe	20			fonctions narratives proppiennes
Bod et al. (2012)	Anglais	4			fonctions narratives et assignation des personnages aux sept catégories proppiennes
Alm et al. (2005)	Anglais	22	1580		émotions et valence
GV-LEx	Français	89	5438	66922	segmentation en épisodes, citations directes, énumérations, attribution des citations, référencement des mentions des personnages, entités nommées

Plus récemment, la constitution de corpus de contes annotés a été relancée par différents enjeux académiques ou industriels. Ces corpus sont généralement conçus pour permettre la mise au point et l’évaluation de procédures d’analyse automatique de texte. Parmi les applications dont la réalisation nécessite d’extraire automatiquement des informations dans des textes assimilables à des contes, on peut mentionner les générateurs d’histoires (Gervás et al., 2005), les applications de narration interactive (Grasbon and Braun, 2001), les robots destinés à interagir avec des enfants (Maarouf and Villaneau, 2012; Mutlu et al., 2006) ou les systèmes de synthèse de parole (Zhang et al., 2003; Mamede and Chaleira, 2004; Alm et al., 2005; Theune et al., 2006; Francisco et al., 2011). Une liste de corpus de contes annotés décrits dans la littérature est fournie dans la table 2.1. On peut y distinguer un certain nombre de tendances.

Ainsi, certains corpus (Malec, 2010; Bod et al., 2012) sont annotés en utilisant le formalisme structuraliste défini par Vladimir Propp (Propp, 1928). Le corpus décrit par Malec (2010) contient 20 contes annotés avec les fonctions narratives de Propp et a été utilisé pour entraîner un système d'annotation automatique. L'utilité d'un tel système est principalement liée à des applications de génération d'histoires (Gervás et al., 2005) ou à la gestion d'histoires interactives (Grasbon and Braun, 2001) dont l'architecture est basée sur le formalisme de Propp. Cette volonté d'extraire des informations structurelles a donné lieu à la mise au point d'un format d'annotation basé sur le XML (Extensible Markup Language) et conçu pour représenter les séquences de fonctions narratives : le PftML (Proppian Fairy Tale Markup Language, Malec, 2004). Ce format a, par la suite, été enrichi pour y intégrer des informations liées à l'analyse linguistique des textes (Lendvai et al., 2010) dans le but d'affiner le schéma d'annotation initialement proposé par Propp. Plus récemment, Bod et al. (2012) décrivent l'accord inter-annotateurs observé pour l'annotation de quatre contes pour lesquels il existe une annotation de référence réalisée par Propp lui-même. La première expérience comporte neuf annotateurs chargés de déterminer les catégories de personnages et les séquences de fonctions narratives. Dans la deuxième expérience, six annotateurs, à qui l'on fournit les catégories de personnages de référence, doivent annoter des séquences de fonctions narratives. Cette étude tend à montrer un faible accord inter-annotateurs dans l'assignement des personnages à l'une des sept catégories définies par Propp. Ces confusions se répercutent sur l'accord entre séquences de fonctions narratives. Même lorsque les catégories des personnages sont fournies aux annotateurs, les auteurs observent une certaine disparité dans l'association des contes à des séquences de fonctions narratives et concluent que la comparaison de ces séquences n'est pas triviale. Les conclusions de cette étude tendent à montrer que les consignes d'annotations définies par Propp ne sont pas assez strictes pour être reproductibles de manière fiable à l'aide de systèmes automatiques.

Les approches cherchant à extraire automatiquement des informations utiles à la synthèse de parole ont principalement consisté à piloter des systèmes de synthèse par sélection d'unités. Dans ce cadre, les informations extraites sont destinées à sélectionner la base d'unités à utiliser pour la synthèse. Chaque base peut contenir des unités correspondant à un locuteur particulier ou des unités correspondant à une émotion ou à une qualité de voix donnée. Le but des systèmes de pilotage automatique de synthétiseurs de parole consiste donc à sélectionner la voix la plus appropriée parmi un ensemble de possibles. Les différents corpus recensés contiennent des annotations qui facilitent le pilotage de systèmes de synthèse parole. L'approche proposée par Zhang et al. (2003) et Mamede and Chaleira (2004) consiste à collecter des corpus contenant des annotations relatives à l'identification et à l'attribution des citations directes aux différents personnages. Ces corpus servent de base aux systèmes de pilotages semi-automatiques ; l'attribution d'une voix particulière à un personnage devant être réalisée manuellement. Une piste d'enrichissement de ces corpus pourrait constituer en l'ajout de méta-informations associées aux personnages qui permettraient un choix automatique de la base de synthèse associée (par exemple voix d'homme, de femme, d'enfant, "d'animal", de "gentil", de "méchant", etc.). Cette approche

montre plusieurs limites. Lorsqu’un conteur raconte une histoire, il incarne tous les personnages – et s’il utilise des stratégies pour différencier vocalement ces personnages, sa voix gardera toujours des caractéristiques propres. En conséquence, une stratégie visant à sélectionner des unités enregistrées par des locuteurs différents en fonction du personnage à incarner ne reproduira pas la situation d’un conteur face à un auditoire, même si le résultat final peut se révéler autant, voir plus captivant. Par ailleurs, les méta-informations associées à un personnage ne décrivent pas la variation de ses affects au cours d’une histoire. Un même personnage peut se montrer agressif, heureux ou méfiant selon les différentes situations auxquelles il est confronté.

Deux corpus de contes conçus pour générer de la parole de synthèse proposent d’associer chaque phrase rencontrée dans les contes à une émotion. Alm et al. (2005) associent chaque phrase à une émotion dite de base : neutre, colère, dégoût, peur, gaieté, tristesse, surprise positive, surprise négative. Ces émotions sont annotées pour chaque conte par deux annotateurs. L’accord inter-annotateurs associé est mesuré sur les différents contes à l’aide du Kappa de Cohen (Cohen et al., 1960) et varie entre 0.24 (accord faible) et 0.51 (accord modéré).

Le corpus Emotales (Francisco et al., 2011) propose deux protocoles d’annotation émotionnelle distincts : le premier consiste à choisir une étiquette émotionnelle parmi 119 propositions, regroupées sous forme d’une ontologie. Un total de 36 annotateurs ont contribué à l’annotation d’Emotales ; chaque conte étant annoté par sept à douze annotateurs différents. Les accords inter-annotateurs sont estimés à l’aide du Kappa de Cohen, en considérant les catégories initiales ou les catégories d’une profondeur donnée dans l’ontologie. Selon le niveau ontologique considéré, le Kappa moyen varie de 0.16 à 0.27, ce qui est considéré comme un accord très faible à faible. Le second protocole consiste à déterminer les émotions exprimées dans chaque phrase dans un espace tridimensionnel selon les dimensions de valence (négative, positive), d’activation (faible, forte) et de contrôle (dominé, dominant). Vingt-six annotateurs ont participé à ce protocole et chaque conte est annoté par six à quatorze annotateurs différents. L’accord est estimé à l’aide de l’écart type observé sur les différentes dimensions ; écart qui est de 0.68 pour la valence, 0.90 pour l’activation et 0.55 pour le contrôle. Ces valeurs sont considérées par les auteurs comme un accord fort et les poussent à préconiser l’emploi de dimensions émotionnelles plutôt que des catégories émotionnelles dans le cadre de l’annotation d’un corpus.

Parmi les autres corpus de contes annotés, les travaux de Goh et al. (2012) proposent d’identifier les personnages principaux de différents contes. Le but du corpus est d’améliorer les capacités des extracteurs d’entités nommées, généralement évalués sur des textes médicaux ou journalistiques. Les auteurs considèrent que les textes de fiction possèdent des propriétés différentes dont l’étude pourrait permettre d’améliorer les performances des systèmes d’extraction d’information pour ce genre de textes.

Le corpus FTC (Fairy Tales Corpus) regroupe ainsi des histoires écrites par des adultes et des enfants (Maarouf and Villaneau, 2012). Il a été réalisé dans le but de modéliser la langue destinée aux enfants : travail nécessaire pour concevoir des interfaces robotiques interagissant avec ce

public. Le corpus est annoté avec des lemmes et des parties du discours corrigés manuellement, et spécifiant le rôle sémantique des verbes. Les noms propres et les pronoms sont référencés, et associés à une catégorie référentielle (prince, éléphant, chaise). Les catégories référentielles sont organisées en quarante catégories ontologiques générales (humain, créature imaginaire, lieu, partie du corps, nourriture...).

À notre connaissance, les corpus GV-LEx et FTC sont les seules ressources de contes annotés disponibles en langue française. Ces deux corpus fournissent des annotations permettant de référencer les différents personnages. Le corpus GV-LEx se distingue en proposant des annotations liées à la structure des contes considérés et aux citations directes, ainsi que des définitions étendues des entités nommées et des énumérations, tandis que le corpus FTC propose des informations linguistiques corrigées à la main (lemmes et parties du discours), l'assignation des personnages à des catégories ontologiques, et l'annotation sémantique du rôle des verbes.

L'annotation des citations directes semble avoir été réalisée à plus grande échelle dans le corpus GV-LEx que dans les autres corpus de contes rencontrés dans la littérature, ce qui en fait une ressource précieuse pour l'entraînement de modèles statistiques. Le schéma d'annotations structurelles proposé reprend un certain nombre des idées introduites par Propp (1928), simplifiées afin d'être utilisées sur des contes d'origines culturelles variées. Les annotations relatives aux entités nommées peuvent être utilisées dans le cadre de l'entraînement de modèles statistiques et de l'évaluation de systèmes automatiques adaptés aux textes de fiction, travail dont la nécessité est mise en évidence par les travaux de Gervás (2010); Goh et al. (2012).

## 2.3 Constitution du corpus GV-LEx

### 2.3.1 Collecte de contes

Les critères de sélection des contes consistaient à choisir des textes pouvant être lus en cinq minutes environ devant un public composé d'enfants âgés de 7 ou 8 ans. En conséquence, le lexique devait être suffisamment simple pour être compris par le public visé. Afin de pouvoir travailler sur la parole des personnages, chaque conte devait contenir un minimum de deux personnages associés à des citations directes. Un total de 89 contes a été retenu : 86 d'entre eux ont été obtenus via un site collaboratif de contes pour enfants libres de droits<sup>1</sup>. Les trois contes restants ont été fournis à titre gracieux par leur auteur : Rose-Marie Vassalo. Les contes retenus sont composés de textes traditionnels ou contemporains et proviennent d'origines culturelles variées. Le corpus collecté contient des contes de diverses catégories : des contes de fées, des contes d'animaux, des contes réalistes, des contes à structure répétitive, etc.

---

1. <http://www.contes.biz>

TABLE 2.2 – Nombre moyen, minimum et maximum d’éléments observés sur les 89 contes suite au processus de tokenisation

	Moyenne	Min	Max	Total
Nb mots	752.1	309	1054	66935
Nb ponctuations	149.7	71	286	13320
Nb phrases	61.2	26	131	5444
Nb paragraphes	19.6	3	41	1742
taille moyenne des phrases (nb mots)	13.0	6.8	22.1	
taille maximum des phrases (nb mots)	40.2	22	116	

### 2.3.2 Normalisation des textes

Les contes du site collaboratif ont été obtenus au format HTML. La conversion des fichiers HTML vers un format en texte seul a été réalisée à l’aide de l’outil Wmatch (Galibert, 2009; Rosset et al., 2009). Un certain nombre d’instructions relatives à la mise en pages ont été conservées dans les fichiers texte résultants. Ainsi, les retours à la ligne et les balises de paragraphe ont respectivement été associés à un ou deux caractères de retour à la ligne. Les fautes d’orthographe et les erreurs grammaticales ont par la suite été corrigées manuellement.

Le logiciel de *tokenisation* du LIMSI (Adda et al., 1997) a été utilisé dans un second temps pour segmenter les textes en unités d’analyse élémentaire (*tokens*). Ces unités correspondent à des mots, des signes de ponctuation, des phrases et des paragraphes. L’utilisation de ce logiciel de tokenisation a été motivée par les traitements à réaliser ultérieurement sur le corpus, tels que l’annotation automatique des parties du discours ou l’alignement semi-automatique entre texte et parole. Les fichiers résultant de cette étape de tokenisation sont tels que les mots et symboles de ponctuation sont séparés par des espaces, les phrases par un caractère de retour à la ligne et les paragraphes par deux caractères de retour à la ligne. Un autre rôle de ce logiciel consiste à désambiguïser les caractères non-alphabétiques. Dans le cas du signe - il faut déterminer s’il s’agit d’un tiret, signe de ponctuation, ou d’un trait d’union. Les mots composés (comme “chauve-souris”) sont associés à un seul *token*. Les clitiques sont séparés de leur mot hôte pour constituer deux tokens distincts (par exemple : “j’ai” est associé aux tokens <j> et <ai>). Les majuscules sont conservées uniquement dans le cas des éléments lexicaux identifiés comme étant des noms propres.

Les résultats du comptage des différents *tokens* obtenus pour l’ensemble des contes suite à ce processus sont présentés dans la table 2.2. Les contes du corpus contiennent une moyenne de 752 mots, 61 phrases et 20 paragraphes. Le premier paragraphe est systématiquement associé au titre du conte considéré. Les stratégies d’indentation (notamment en ce qui concerne l’utilisation des retours à la ligne) sont très variables selon les auteurs. Le nombre de paragraphes varie (si l’on exclut le titre) de 2 à 39 selon les contes. Ceux contenant un grand nombre de paragraphes



sont généralement associés à des stratégies d'indentation consistant à insérer un marqueur de paragraphe avant chaque citation directe. Les phrases les plus courtes sont généralement associées aux citations directes et aux motifs précisant les citations (cf. exemple 1). Les phrases les plus longues ont été observées dans le cadre de descriptions et dans les contes à structure répétitive (cf. exemple 2).

(1) “- Oui !

dit la souris.”

(2) “Viens ronger le cordon, qui refuse d'étrangler le forgeron, qui refuse de briser cette lame, qui refuse de tuer le taureau, qui refuse de boire l'eau, qui refuse d'éteindre ce feu, qui refuse de brûler ce bâton, qui refuse de frapper Brirouch, qui refuse de dîner, parce qu'il a perdu son chevreau.”

### 2.3.3 Protocole d'annotation

La tâche d'annotation des 89 contes a été confiée à l'entreprise Syllabs. Deux annotateurs (familiers avec les tâches d'annotation linguistique) ci-après désignés comme A1 et A2, ont respectivement annoté 61 et 35 contes, pour un total de 96 annotations distinctes. Sept contes ont donc été annotés par A1 et A2 et seront utilisés par la suite pour estimer l'accord inter-annotateur (section 2.6). Un schéma de stockage XML<sup>2</sup> hiérarchique, associé à une grammaire de classe de document (DTD<sup>3</sup>) a été utilisé pour encoder les annotations. Des consignes d'annotation et un exemple de conte entièrement annoté ont été fournis aux annotateurs. Les textes ont été annotés à l'aide d'un logiciel d'annotations multi plateforme développé par Syllabs pour répondre aux besoins spécifiques du projet GV-LEx. Ce logiciel permet de préserver l'indentation du texte normalisé, qu'il s'agisse de retours à la ligne simples ou multiples marquant les paragraphes. Il permet une visualisation intuitive des éléments annotés. Le logiciel utilise la définition de la grammaire du document XML définie par le LIMSI, au format DTD, pour générer les menus et sous-menus associés à un schéma hiérarchique quelconque. Cette caractéristique permet une certaine flexibilité et laisse la possibilité de redéfinir le schéma d'annotation, tout en gardant le même logiciel.

L'interface d'annotation du logiciel est illustrée par la figure 2.1. Le menu de navigation gauche permet de sélectionner un conte dans le corpus. Le menu de la partie supérieure droite renseigne sur le nombre d'annotations contenu dans le conte courant et permet de sélectionner les marqueurs à afficher. Le menu inférieur permet de sélectionner chaque marqueur séparément et d'éditer ses attributs. Les différents niveaux hiérarchiques des annotations correspondants à un mot ou à une zone de texte donnés sont accessibles via des sous-menus. Dans cet exemple, les entités étendues relatives aux personnes sont représentées avec un fond bleu et les citations directes sont affichées avec une police rouge.

---

2. Extensible Markup Language

3. Document Type Definition

## 2.3. Constitution du corpus GV-LEx

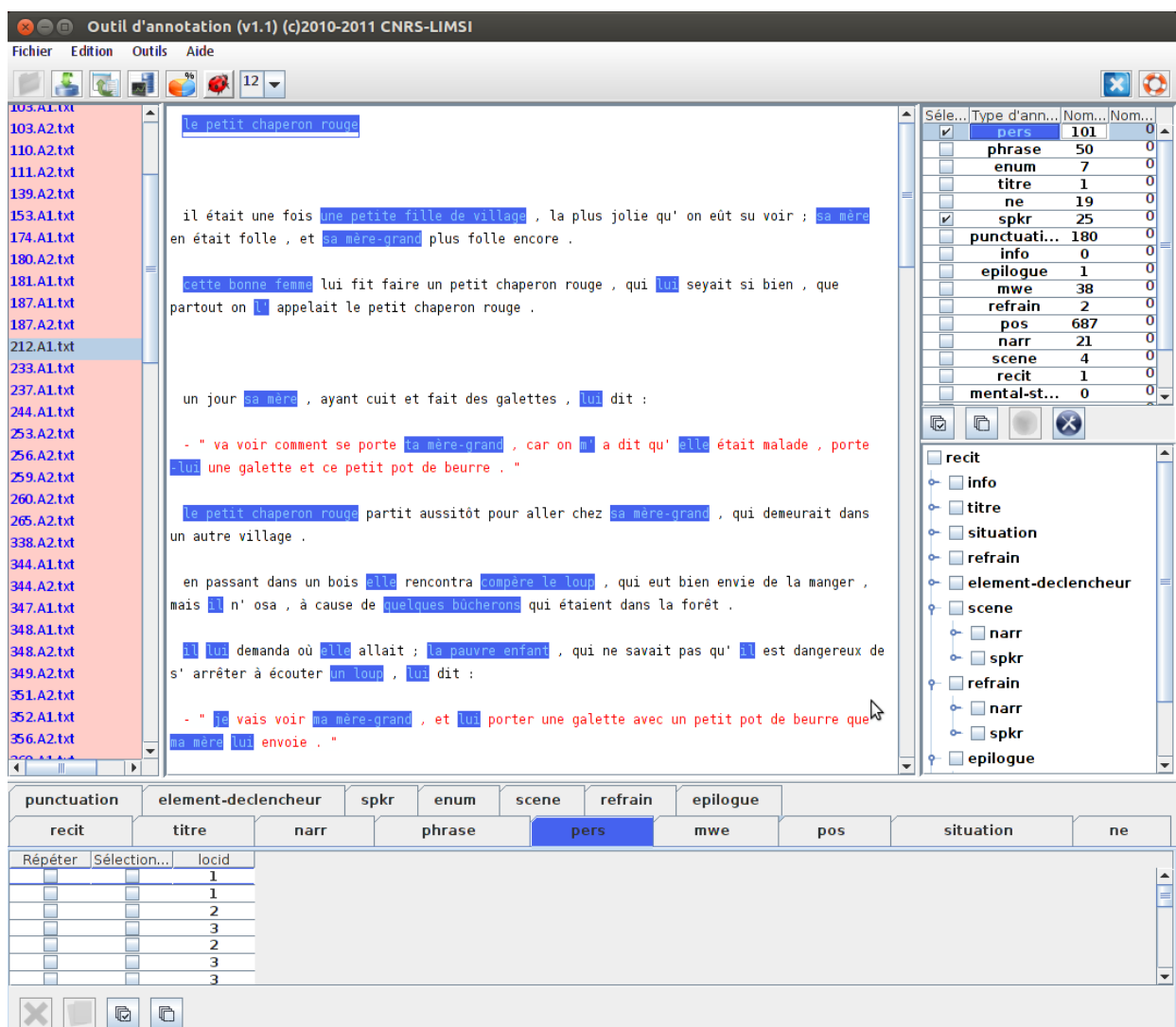


FIGURE 2.1 – Capture d'écran du logiciel d'annotation linguistique conçu par l'entreprise Syllabs pour les besoins spécifiques du projet GV-LEx.

TABLE 2.3 – Marqueurs XML utilisés dans les corpus GV-LEx

Marqueur	Description
Marqueurs Structurels	
<b>gvlex_tale</b>	Contient l'ensemble du conte
<b>title</b>	Le titre de l'histoire
<b>exposition</b>	Introduction, situation initiale et présentation des personnages
<b>triggering-event</b>	Élément déclencheur qui modifie la situation initiale et engendre les péripéties
<b>scene</b>	Unité atomique, passage du récit qui fait avancer l'histoire
<b>refrain</b>	Passage du récit récurrent à quelques détails près
<b>epilogue</b>	Conclusion, dénouement et/ou morale du récit
Tours de Parole	
<b>narr</b>	Tour de parole du narrateur.
<b>spkr</b>	Citation directe d'un personnage du conte, identifié avec l'attribut anaphorique <b>persid</b>
Éléments Lexicaux et semi-structurels	
<b>pseudo_sentence</b>	Pseudo-phrase qui ne peut s'étendre sur plusieurs tours de parole
<b>enum</b>	Énumération étendue
Entités nommées étendues	
<b>person</b>	Mention d'un personnage identifié à l'aide de l'attribut anaphorique <b>persid</b>
<b>place</b>	Mention d'un lieu
<b>time</b>	Mention d'une date ou d'une époque
Éléments Atomiques	
<b>word</b>	Mot contenant les attributs <b>small_pos</b> et <b>large_pos</b> stockant les parties du discours
<b>punctuation</b>	Signe de ponctuation

## 2.4 Schéma d'annotation

Cette partie décrit les schémas d'annotation définis pour le corpus ainsi que les consignes ayant été données aux annotateurs. Les contes sont divisés en *épisodes* permettant de décrire la structure narrative. La distinction entre les citations directes et le reste du texte est réalisée via l'annotation en *tours de parole*. Des définitions étendues des énumérations et des entités nommées décrivant des dates, des lieux ou des personnes sont utilisées. Les tours de parole et les entités relatives aux personnes sont enrichis avec des informations anaphoriques permettant de repérer les mentions des différents personnages et de leur attribuer les citations directes. L'ensemble des marqueurs XML utilisés pour annoter les textes sont décrit dans la table 2.3. La DTD<sup>4</sup> définissant la grammaire des annotations XML, et un exemple de conte annoté sont fournis dans les annexes A et B.

---

4. Document Type Definition

### 2.4.1 Énumérations étendues

Les énumérations étendues sont associées à toute structure assimilable à une énumération ou à une liste. Il peut s’agir d’une énumération d’objets (cf. exemple 3), d’adjectifs (cf. exemple 4), d’actions, comme de personnages.

(3) Suzon ouvrit ensuite tout grand <enum>les portes, les fenêtres et même les volets</enum>.

(4) Vous êtes <enum>si belle, si bonne et si honnête</enum>, que je ne puis m’empêcher de vous faire un don.

### 2.4.2 Entités nommées étendues

Les entités nommées relatives aux personnes sont définies pour identifier toutes les mentions dans un conte d’un personnage ou d’un groupe de personnages. Ces annotations sont enrichies avec des informations anaphoriques qui permettent de distinguer les différents personnages. Les mentions des personnages peuvent être constituées par des noms propres, des pronoms ou tout groupe nominal significatif. Les pronoms possessifs isolés ne sont pas considérés comme des mentions explicites d’un personnage, par contre ils peuvent être utilisés au sein d’un groupe nominal pour faire référence à un personnage (cf. exemple 5).

(5)

<person id=3>Il</person> demanda à <person id=6>la mère de <person id=5>son fils</person></person>, ce qu’<person id=3>il</person> pouvait faire pour qu’<person id=6>elle</person> <person id=3>lui</person> pardonna.

Les contes sont des fictions et les personnages peuvent être des humains, des animaux, des plantes ou des objets. Les entités nommées spatiales marquent toutes les structures se référant à un lieu au sens large. Il peut s’agir d’une ville (“*Paris*”) comme d’un endroit dont l’emplacement est approximatif (par exemple : “*quelque part en Afrique*”), “*là bas*”, “*dans la chambre*”). Les entités nommées temporelles sont utilisées pour identifier les dates précises, les périodes ou les expressions temporelles relatives (par exemple : “*un an plus tard*”, “*après une courte discussion*”, “*la veille de Noël*”, “*il était une fois*”). Ces entités nommées peuvent être définies de manière hiérarchique ou à l’aide d’opérations de composition permettant de décrire une entité à l’aide d’une autre. Dans l’exemple 6, l’entité spatiale “entre les enfants et la vieille” est définie relativement à la position d’autres entités faisant référence à des personnes : “les enfants” et “la vieille”.

(6) <time>Aussitôt</time> une toile d’araignée s’élève <place>entre <person id=5>les enfants</person> et <person id=6>la vieille</person></place>.

### 2.4.3 Tours de parole

L'annotation des textes en tours de parole permet de distinguer les citations directes du reste du texte (exemple 7). Les citations directes rencontrées dans le corpus ne sont pas systématiquement associées à des signes de ponctuation tels que les guillemets ou les tirets. Les tours de parole des personnages, annotés à l'aide de balises `<spkr>`, correspondent aux portions de texte assimilables à une citation directe d'un personnage. Un tour de parole est défini comme une zone contigüe de texte attribué à un même personnage. La taille de cette région peut varier de quelques mots à quelques phrases. Les balises entourant cette région contiennent un attribut anaphorique faisant référence au personnage à qui est attribué la citation. Il est ainsi possible que le narrateur se cite lui-même (cf. exemple 8), dans le cas des récits au cours desquels le narrateur raconte des événements auxquels il affirme avoir participé. Par convention, l'attribut anaphorique associé au narrateur est 0. Lorsque la portion de texte considérée n'est pas une citation directe, on considère qu'il s'agit du tour de parole du narrateur, entouré de balises `<narr>`.

(7) `<spkr id=2>"eh bien!",</spkr> <narr>dit <person id=2>le loup</person>  
,</narr> <spkr id=2>"<person id=2>je</person> veux y aller voir aussi."</spkr>`

(8) `<spkr id=0>-<person id=0>je</person> crois que ça doit être un hibou,</spkr>  
<narr>répondis-<person id=0>je</person> sans trop y croire.</narr>`

### 2.4.4 Structure narrative

Les annotations relatives à la structure narrative consistent à segmenter les contes de manière à les représenter sous forme de séquence d'*épisodes*. Les annotations structurelles définies pour l'annotation du corpus sont inspirées par les *épisodes* (van Dijk (1982)), les *fonctions narratives* (Propp (1928)) et les *sous-thèmes* (Hearst (1997)). Il s'agit de segments linéaires sans chevauchement, pouvant s'étendre sur des régions composées d'une ou plusieurs phrases normalisées. Les frontières des épisodes sont généralement associées à des changements temporels, spatiaux ; des changements de sujets ou à l'introduction de nouveaux personnages dans le récit. Les annotations structurelles ne sont pas nécessairement associées aux marques de paragraphe rencontrées dans le texte. En effet, il arrive que des marques de paragraphe ne soient insérées que pour aérer l'aspect physique d'un texte, sans pour autant être associées à une information linguistique particulière (Hearst, 1997).

Les marques de paragraphes ont été conservées dans les textes présentés aux annotateurs. Un ensemble de cinq catégories sont définies pour caractériser les épisodes rencontrés : *situation initiale*, *élément déclencheur*, *scène*, *refrain*, *épilogue*. Les catégories ont été définies de manière suffisamment générale pour décrire des contes ayant été écrits à des époques différentes ou ayant

des origines géographiques variées. La *situation initiale* correspond généralement aux premières phrases du récit. Elle décrit le contexte du récit, et présente l’histoire et les personnages. On peut l’assimiler à la notion de situation initiale définie par Propp (Propp, 1928). L’*élément déclencheur* altère la situation initiale et engendre les péripéties. Il peut s’agir de la mort d’un personnage, de la violation d’une interdiction. Par analogie avec la théorie proppienne, cet épisode s’étend de la fin de la situation initiale, jusqu’au départ du héros. Une *scène* est un passage du récit qui fait avancer l’histoire. Il s’agit de la catégorie épisodique la plus fréquente. La décision de changement de scène s’appuie essentiellement sur des marqueurs thématiques et structuraux (apparition d’un nouveau personnage, changement spatial ou temporel, nouveau paragraphe) ou alors sur l’identification d’un refrain. Un *refrain* est un passage du récit récurrent dans la forme et, à quelques détails près, dans le contenu. Il peut s’agir de phrases ou de paragraphes strictement identiques qui se répètent régulièrement au cours du récit, ou de portions de texte qui subissent quelques modifications de contenu. L’*épilogue* représente généralement les dernières lignes d’un récit. Il s’agit de la conclusion, du dénouement, de la morale du récit. Dans la théorie proppienne, cet épisode inclut la reconnaissance du héros, la punition du méchant, la transfiguration du héros et le mariage.

L’ordre dans lequel les différents types d’épisodes peuvent être rencontré est défini dans la DTD présentée dans l’annexe A. Un conte commence par un titre, la présence de tous les autres marqueurs structurels est facultative. Dans le cas où les marqueurs sont identifiés, on trouvera d’abord la situation initiale, puis l’élément déclencheur, ensuite des scènes, et enfin un épilogue. Toutes ces structures peuvent être entrecoupées de refrains.

## 2.5 Couverture et Répartition des Annotations

### 2.5.1 Unités textuelles

Cette section décrit la répartition des annotations dans les 89 contes du corpus textuel. Les annotations sont décrites en terme de nombre d'unités annotées et de pourcentage de couverture sur le corpus. Ce pourcentage de couverture est défini comme le pourcentage de mots du corpus inclus dans une unité d'un type donné, en excluant de la mesure les marqueurs de ponctuation.

Les mesures sont effectuées en considérant un seul annotateur par conte afin d'éviter de sur-représenter les contes ayant été annotés deux fois. Seules les annotations réalisées par l'annotateur A2 ont été prises en compte dans le cas des contes annotés deux fois, afin d'équilibrer le nombre d'annotations considérées par annotateur. En conséquence, les mesures réalisées dans cette partie sont obtenues à partir de 54 annotations réalisées par A1 et 35 annotations par A2.

Un résumé des distributions d'annotations et des pourcentages de couverture est fourni par la table 2.4. Les tailles médiane, moyenne et maximum de chaque annotation sont exprimées en nombre de mots couverts par une annotation donnée.

Un conte contient toujours un titre et 5 scènes en moyenne. La majorité des contes contient une situation initiale, un épilogue, et dans une moindre mesure un événement déclencheur. Les scènes sont en général plus longues que les autres catégories d'épisodes et couvrent plus de 70% du corpus. Les épilogues et les événements déclencheurs sont en général plus courts que les situations initiales. Un cinquième des contes annotés contient des refrains. Il s'agit de la catégorie épisodique montrant la plus faible couverture et la plus petite taille moyenne, à l'exception du titre.

Le corpus contient un grand nombre de citations directes de personnages (les passages du corpus annotés comme faisant partie du tour de parole d'un personnage), tandis que des citations directes du narrateur n'ont été observées que pour un seul conte. En fonction du conte considéré, la couverture des citations directes peut varier de 4,5% à 72%, avec une couverture moyenne aux alentours de 30%. Au total, 381 personnages différents réalisent ces citations directes dans le corpus, avec un nombre médian de 4 personnages par conte et un maximum de 14 personnages différents prenant la parole. Les phrases attribuées au narrateur sont généralement plus longues que les phrases attribuées aux personnages avec une taille médiane respective de 11 vs. 7 mots. Le *texte cité* correspond aux portions de texte comprises entre des guillemets, ou aux phrases normalisées commençant par un tiret et représente 23,7% des mots contenus dans le corpus. Ce *texte cité* est constitué pour 88% par des citations directes et pour 12% par des incises ou des mots entre guillemets. Le *texte cité* est donc constitué en majeure partie de citations directes. Cependant, les citations directes ne se trouvent pas uniquement dans le *texte cité*. On trouve en effet une part non négligeable de citations directes en dehors du texte cité (12,1%). Cela représente presque un tiers de l'ensemble des citations directes observées sur l'ensemble du corpus.

Les annotations relatives aux énumérations couvrent un cinquième du corpus et 18,2% de ces structures couvrent des phrases entières. Les entités nommées étendues relatives aux personnages

TABLE 2.4 – Distribution des annotations dans les 89 contes du corpus texte. Le pourcentage de couverture correspond au pourcentage de mots du corpus inclus dans une unité d'un type donné, en excluant de la mesure les marqueurs de ponctuation

Annotation	Nombre	Couverture (%)	taille médiane	taille moyenne	taille maximum
Épisodes					
Titre	89	0.6	4	4.6	9
Situation initiale	88	11.6	64	88.5	483
Événement déclencheur	83	9.5	39	76.3	372
Scène	443	70.7	80	106.8	599
Refrain	75	2.8	20	25.0	109
Épilogue	88	4.8	32	36.4	95
Tours de parole du narrateur et des personnages					
Personnage	1509	30.1	8	13.4	148
Narrateur	1205	69.9	15	38.8	635
Éléments semi-structuraux et lexicaux					
Énumération	1270	20.0	9	10.5	86
Entité nommée spatiale	1377	6.8	3	3.3	13
Entité nommée temporelle	910	3.5	2	2.6	14
Entité nommée personne	9500	22.3	1	1.6	18



sont beaucoup plus fréquentes que les entités nommées temporelles et spatiales : avec en moyenne 107 entités de personnage par récit, contre respectivement 15 et 10. Un certain nombre d'entités nommées composées ont été observées dans le corpus : il s'agit d'entités définies relativement à d'autres entités, via des mécanismes hiérarchiques. L'analyse des compositions d'entités montre que 13,6% des entités spatiales et 2% des entités temporelles sont définies relativement à une entité de personne (par exemple : “la maison de Monsieur Martin” est une entité spatiale, définie relativement à l'entité personne “Monsieur Martin”). Les autres schémas de composition sont beaucoup moins fréquents et concernent moins d'un pourcent des cas rencontrés (par exemple une personne définie relativement à une autre personne, ou une personne définie relativement à un lieu, etc.).

Les propriétés syntaxiques des entités de personne sont décrites à l'aide de parties du discours obtenues de manière automatique. Un modèle d'annotation automatique basé sur l'algorithme TreeTagger (Schmid, 1994) est utilisé. Les conditions d'apprentissage du modèle sur la langue française sont détaillées dans Stein and Schmid (1995). Les entités de personnes sont principalement associées à des pronoms personnels (52%), à des articles suivis de noms (15%), à des noms propres (7%), à des pronoms possessifs suivis de noms (3,6%) et à des structures constituées d'un article, d'un adjectif puis d'un nom (2,6%). Les 17% restants ont des structures et des complexités variées. Les entités de personnes étendues constituent pour 99.2% des structures de 5 mots au maximum.

### 2.5.2 Annotations anaphoriques et chaînes de coréférence

L'étude de la couverture des annotations anaphoriques est réalisée en regroupant les unités faisant référence à une même entité sous forme de *chaîne de coréférence*. Une *chaîne de coréférence* est définie comme la liste des unités annotées au sein d'un même document partageant le même attribut anaphorique. La liste est ordonnée selon l'ordre d'apparition des unités au sein d'un document, et peut contenir des doublons. Dans le cas des entités de personne (section 2.4.2), la chaîne contient toutes les unités faisant référence à un même personnage (exemple : “le petit chaperon rouge”, “un petite fille”, “lui”, “l' ”, “elle”, “la pauvre enfant”). Dans le cas des tours de parole (section 2.4.3), les éléments de la liste correspondent aux tours de parole attribués un même personnage. Chaque élément correspond à une pseudo-phrase, ou à plusieurs pseudo-phrases adjacentes.

La table 2.5 fournit une description des chaînes de coréférence relatives aux tours de parole et aux entités de personnes. Les chaînes considérées sont associées aux tours de parole (**TP**), aux entités de personne citées directement au moins une fois dans le récit (**EPP** : Entités Personne Parlante), et aux entités de personnes qui ne sont pas citées directement dans le récit (**EPNP** : Entités Personne Non Parlante). Pour chacune de ces catégories la table décrit le nombre de chaînes observées ainsi que le nombre d'unités associées. Viennent ensuite les pourcentages de chaînes à un seul élément qui correspondent soit aux tours de parole associés à un personnage cité directement une seule fois au cours du récit, soit à un personnage mentionné une seule fois dans

TABLE 2.5 – Description des chaines de coréférence associées aux tours de parole (TP), aux entités personne parlante (EPP) et aux entités personne non parlante (EPNP)

Type de chaîne	Nb. chaines	Nb. unités	Pourcentage de chaines à un seul élément	Longueur médiane	Troisième quartile	Longueur maximale
TP	381	1509	29.9%	3	5	20
EPP	377	7287	8.8%	13	28	90
EPNP	545	2213	40.4%	2	4	71

le récit. La distribution des longueurs de chaines de coréférence est décrite à l’aide de la médiane, du troisième quartile, et de la valeur maximale. Le nombre d’unités observées est moins important pour les tours de parole, car ils couvrent de plus grandes zones de texte. Les 9500 entités nommées étendues relatives aux personnages font référence à un total de 922 personnages distincts. 41% des personnages identifiés dans les contes sont associés à au moins un tour de parole. Une grande partie des marqueurs d’entités de personne (76,7%) fait référence à des personnages auxquels est associé au moins un tour de parole. Plus d’un quart (27%) des entités de personne font référence à des personnages qui ne sont mentionnés qu’une seule fois dans tout le récit. La taille médiane du nombre de références portant sur les personnages associés à au moins un tour de parole est de 13, le nombre de références maximales pour un personnage étant de 90. La taille des chaines de coréférence associées au tour de parole décrit le nombre de zones de texte contigües attribuées à un même personnage. Ces zones peuvent aussi bien contenir quelques mots que quelques phrases.

## 2.6 Accord inter-annotateurs

L'annotation d'un corpus est une tâche d'interprétation qui comporte nécessairement une part de subjectivité (Leech et al., 1997). Les désaccords peuvent provenir d'erreurs d'annotation, de consignes ambiguës ou de tâches trop complexes. Plusieurs interprétations peuvent être proposées pour décrire un même motif et il n'est pas toujours possible de déterminer laquelle est la plus pertinente.

L'analyse des accords inter-annotateurs permet d'estimer la robustesse d'un schéma d'annotation. Un score élevé d'accord entre annotateurs permet de considérer des annotations comme fiables et reproductibles. Un faible accord peut donner lieu à des reformulations de consignes d'annotation jusqu'à obtention d'un accord jugé suffisant (Bittar et al., 2012). Dans le cas d'annotations destinées à entraîner et à évaluer des systèmes de traitement automatique, la connaissance de l'accord inter-annotateurs fournit une information sur les plafonds de performance pouvant être attendus : on peut difficilement espérer que l'accord entre des annotations manuelles et les prédictions d'un système automatique soit meilleur qu'entre deux annotateurs entraînés. Comme nous l'avons signalé, les mesures d'accord présentées dans cette section ont été obtenues à partir des 7 contes annotés par les annotateurs A1 et A2.

### 2.6.1 Segmentation en épisodes

La segmentation du texte en épisodes étiquetés combine la segmentation du texte en unités et le choix d'une étiquette décrivant son contenu. L'unité atomique choisie pour mesurer la fiabilité de cette tâche est la phrase normalisée. Tous les contes du corpus texte commencent par un titre correspondant à la première ligne du fichier et le titre est systématiquement séparé du reste du conte par des marqueurs de paragraphe. Pour éviter de prendre en compte cette tâche triviale dans les mesures d'accord, l'annotation des titres a été exclue des mesures présentées ci-dessous.

Afin d'annoter les différences entre annotateurs, la procédure suivante est utilisée. Chaque phrase normalisée est associée à une étiquette. S'il s'agit de la première phrase d'un épisode, l'étiquette attribuée est le nom de l'épisode. Dans le cas contraire, l'étiquette attribuée est "pas de frontière". Par exemple, si l'on se rapporte à la table 2.6, qui contient la matrice de confusion inter-annotateurs pour la segmentation en épisodes, on observe que la valeur 7 trouvée à la ligne *scène* et à la colonne *pas de frontière* indique que 7 phrases ont été détectées par A1 comme des frontières de scènes, tandis que A2 les situe à l'intérieur d'un épisode. La taille moyenne des segments annotés est de 6,95 phrases pour l'annotateur A1 et de 4,91 pour A2. L'accord maximal observé correspond à l'annotation des refrains. L'avis des annotateurs varie selon la présence et les bornes de l'introduction et de l'élément déclencheur. Les annotateurs s'accordent sur la localisation du commencement de l'épilogue pour 5 contes sur 7. Bien qu'un accord substantiel soit observé pour les frontières des scènes, on observe une tendance de la part d'A2 à considérer de plus petites scènes, ce qui introduit des différences.

TABLE 2.6 – Matrice de confusion entre A1 et A2 pour la segmentation en épisodes.

A1\A2	Intro.	Evt. déclencheur	Scène	Épilogue	Refrain	Pas de frontière
Intro.	6	0	0	0	0	0
Evt. déclencheur	1	2	1	0	0	2
Scène	0	1	25	0	0	7
Épilogue	0	0	1	5	0	1
Refrain	0	0	0	0	8	0
Pas de frontière	0	4	29	2	0	322

TABLE 2.7 – Table d’interprétation du Kappa de Cohen

$\kappa$	Interprétation
$< 0$	Désaccord
0.0 0.20	Accord très faible
0.21 0.40	Accord faible
0.41 0.60	Accord modéré
0.61 0.80	Accord fort
0.81 1.00	Accord presque parfait

Pour la suite des observations, l’accord est mesuré sur la position des frontières uniquement et non plus sur les étiquettes des épisodes. Ce choix est fait pour se concentrer sur les frontières déterminées par les annotateurs, et non pas sur leur interprétation des étiquettes des différentes sortes d’épisodes. Pour cela, les contes sont préalablement convertis en une séquence de valeurs binaires, dans laquelle chaque élément correspond à une frontière de phrase. L’élément prend la valeur de 1 si la frontière correspond à une frontière d’épisode et 0 sinon (observation négative). Un conte contenant  $N$  épisodes aura en conséquences  $N-1$  frontières d’épisodes. Pour éviter de prendre en compte l’accord sur des cas triviaux, la frontière entre le titre et le début du conte n’est pas prise en compte dans ces représentations binaires. Pour la suite de cette analyse, on considèrera pour chaque annotateur la chaîne binaire correspondant à la concaténation des représentations binaires des sept contes annotés. Les estimations d’accord qui en découlent sont rapportées dans la table 2.8.

Les différentes métriques permettant de mesurer un accord inter-annotateurs sont destinées à traiter différentes situations, et donnent donc des résultats différents sur ces mêmes données - indiquant plusieurs types de variations entre annotateurs. Le Kappa de Cohen (Cohen et al., 1960) constitue la seule mesure utilisée dans cette évaluation prenant en compte l’accord par hasard. Il est défini par l’équation 2.1 présenté ci-dessous.

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \quad (2.1)$$

Tel que  $\Pr(a)$  soit l’accord relatif entre codeurs et  $\Pr(e)$  la probabilité d’un accord aléatoire. Il peut être interprété à l’aide de la table de correspondance 2.7.

TABLE 2.8 – Accord sur la localisation des frontières d'épisodes entre A1 et A2, estimé à l'aide des métriques Kappa, F-Measure, WindowDiff, Weighted WindowDiff et Pk

Kappa	F-Measure	WD	WWD	Pk
0.59	0.66	0.32	0.36	0.28

Un Kappa de 0,59 est observé entre les deux annotateurs, ce qui est compris dans la plage des accords observés sur des tâches similaires (Artstein and Poesio, 2008). La F-mesure (définie dans l'équation 2.4 de la section 2.6.2) peut être considérée comme un estimateur d'accord similaire au Kappa de Cohen dans des cas de figure où le nombre d'observations négatives est élevé (Hripcsak and Rothschild, 2005). Dans notre cas, le nombre de frontières de phrases qui ne sont pas des frontières d'épisodes est environ 6 fois plus important que les autres frontières ; l'usage de la F-Mesure semble donc pertinent.

Le Kappa et la F-Mesure ont pour particularité de comptabiliser indifféremment les erreurs de frontière quelle que soit leur distance avec la frontière de référence. Les estimations d'accord liées à ces mesures peuvent être considérées comme pessimistes : les annotateurs peuvent avoir tendance à s'accorder sur le cœur des segments et à être en désaccord sur la localisation des frontières. Les métriques Pk (Beeferman et al., 1999) et WindowDiff (*WD*) (Pevzner and Hearst, 2002) (cf. équations 2.2 et 2.3) pénalisent les erreurs de frontière proportionnellement à leur distance par rapport aux frontières de référence. Ces métriques sont initialement conçues pour estimer la qualité d'une segmentation obtenue via un système automatique en la comparant à une segmentation de référence, jugée fiable. Leur utilisation pour estimer l'accord inter-annotateur a été proposée par Artstein and Poesio (2008). Ces deux mesures consistent à comparer les segmentations par le biais de fenêtres s'étendant sur plusieurs unités et à retourner un score de dissimilitude. Un score de 0 correspond au score obtenu pour deux chaînes identiques. Ces mesures ont été utilisées de manière complémentaire. Pk est connue pour pénaliser davantage les faux négatifs que les faux positifs (Pevzner and Hearst, 2002), tandis que WindowDiff a tendance à favoriser les segmentations contenant moins de frontières (Lamprier et al., 2007).

$$PK(r, h) = \frac{1}{N - K} \sum_{i=1}^{N-k} (|\delta(r_i, r_{i+k}) - \delta(h_i, h_{i+k})|) \quad (2.2)$$

$$WD(r, h) = \frac{1}{N - K} \sum_{i=1}^{N-k} (|b(r_i, r_{i+k}) - b(h_i, h_{i+k})| > 0) \quad (2.3)$$

La variable  $r$  correspond à la segmentation de référence,  $h$  à la segmentation testée,  $k$  à la taille de fenêtre,  $b(i, j)$  au nombre de frontières de segments entre la  $i$ ème et la  $j$ ème frontière de phrase rencontrée dans le récit.  $\delta(i, k)$  retourne 1 s'il y a au moins une frontière de segment entre la  $i$ ème et la  $j$ ème frontière et 0 sinon.  $N$  représente le nombre de frontières de phrases rencontrées dans le texte. La valeur recommandée pour la taille de fenêtre  $k$  est la moitié de la taille moyenne des segments rencontrés dans la segmentation de référence. Dans le cadre de cette évaluation, la

longueur moyenne de segment est estimée sur l'ensemble des annotations contenues dans le corpus, en excluant les sept annotations réalisées par A1 pour lesquelles il existe une annotation réalisée par A2. Le choix de prendre en compte l'ensemble du corpus, plutôt que les seuls sept contes sur lesquels est estimé l'accord, a été fait afin d'estimer une taille moyenne de segment (et donc une taille de fenêtre puis des estimations d'accord) comparable aux estimations de performance des systèmes automatiques réalisés dans les chapitres suivants. Un total de 688 frontières d'épisodes est observé parmi les 5260 frontières de phrases contenues dans le corpus. La taille moyenne des segments observés est égale à 7,65 phrases, ce qui donne une taille de fenêtre  $k$  égale à 4.

À ma connaissance, les métriques Pk et WindowDiff n'ont pas encore été utilisées pour décrire un accord inter-annotateurs observé sur des tâches de segmentation. Leur utilisation dans ce cadre reste cependant cohérente avec les recommandations formulées par (Artstein and Poesio, 2008), ainsi qu'avec les études de (Mathet et al., 2012). Les mesures effectuées dans cette section contribuent à fournir des données de comparaisons pour les études à venir. Dans le cadre de ce manuscrit, elles fournissent de plus une base de référence qui sera utilisée pour évaluer les modèles de segmentation en épisodes qui seront présentés par la suite.

### 2.6.2 Identification des éléments structurels et lexicaux

Cette section décrit l'accord observé pour les tâches d'annotation relatives aux unités lexicales que sont les énumérations et les entités nommées spatiales, temporelles et de personne. L'accord pour l'annotation des tours de parole est également traité. Les tours de parole couvrant l'intégralité du texte des contes, chaque mot est attribué soit à un personnage, soit au narrateur. Il est aussi possible de décrire l'accord pour les tours de parole en ignorant les unités correspondant aux tours de parole du narrateur, c'est-à-dire en ne prenant en compte que les tours de parole des personnages comme s'il s'agissait d'unités lexicales standards.

Sur la base des travaux de Grouin et al. (2011) et Fort et al. (2012), un ensemble de mesures ont été utilisées pour évaluer les accords inter-annotateurs pour ces types d'annotations : on utilisera donc la F-Mesure stricte ( $F_S$ ), la *half-point F-Measure* ( $F_H$ ), le *slot error rate strict* ( $SER_S$ ) et le *half-point slot error rate* ( $SER_H$ ). On se référera aux équations 2.4 à 2.7 pour plus de détails. Les accords liés au hasard sont pris en compte en utilisant une adaptation du Kappa de Cohen décrite dans Grouin et al. (2011). Cette adaptation consiste à contourner le problème de modélisation d'un accord aléatoire, en considérant chaque unité annotée au moins une fois comme une proposition d'annotation (indépendamment de sa catégorie). Le Kappa est alors calculé sur la base de chaque réponse de chaque annotateur. Dans les mesures suivantes, un total de 1152 propositions d'annotations est considéré, qui correspond à l'union de toutes les unités identifiées par A1 ou A2 comme des entités ou comme des énumérations. Les tours de parole diffèrent des autres unités considérées car ils tendent à s'étendre sur plusieurs phrases plutôt que sur plusieurs mots et ne peuvent être traités comme des unités lexicales standard. Pour cette raison, l'estimation de l'accord pour les tours de parole n'est pas réalisée à l'aide du Kappa de Cohen. Les propositions

d'annotation correspondantes ne sont pas prises en compte pour estimer l'accord sur les autres unités lexicales.

$$F_S = 100 \frac{2sm}{|A1| + |A2|} \quad (2.4)$$

$$F_H = 100 \frac{2sm + pm}{|A1| + |A2|} \quad (2.5)$$

$$SER_S = 100 \frac{2e + 2pm}{|A1| + |A2|} \quad (2.6)$$

$$SER_H = 100 \frac{2e + pm}{|A1| + |A2|} \quad (2.7)$$

où  $|A1|$  et  $|A2|$  correspondent respectivement au nombre d'unités d'un type donné (entités temporelles, tours de parole, ...) identifiées par A1 ou A2.  $sm$  (strict match) représente le nombre d'unités strictement identiques identifiées par A1 et A2 pour un type d'unité donné.  $pm$  (partial match) représente le nombre d'unités partiellement identiques identifiées par A1 et A2 : couvrant une portion commune de texte, mais ayant des frontières différentes.  $e$  est le nombre d'unités d'un type donné pour lesquelles il n'y a pas de correspondance entre les annotations de A1 et A2 et se trouve défini comme  $e = |A1| + |A2| - 2sm - 2pm$ .

Les estimations d'accord pour les annotations lexicales et structurelles sont rapportées dans la table 2.9. Quelle que soit la métrique considérée, les unités correspondant aux tours de parole sont caractérisées par un accord proche de la perfection. La difficulté spécifique liée à l'annotation des entités de personne est due la nature fictive des contes : une entité de personne peut être utilisée pour désigner n'importe quelle entité considérée comme un personnage agissant, qu'il s'agisse d'un humain, d'une plante, d'un objet ou d'un animal. L'accord observé pour cette tâche reste cependant relativement haut et largement supérieur à l'accord observé pour les autres unités lexicales. Dans le cas des énumérations (et dans une moindre mesure pour les entités spatiales), on observe une importante proportion d'accords partiels. Cela traduit une difficulté des annotateurs à déterminer les frontières exactes des unités annotées, ce qui s'observe dans les plus grands écarts entre variantes strictes et partielles (half-point) des mesures d'estimation d'accord considérées. Malgré l'accord moins important observé sur les entités spatiales et temporelles, les mesures réalisées restent cohérentes avec d'autres tâches d'annotations similaires (Zhang et al., 2010).

### 2.6.3 Référencement des personnages

Cette partie présente les estimations d'accord relatives aux annotations anaphoriques associées aux entités personne, et aux tours de parole (l'accord sur l'identification de ces structures a été présenté séparément dans la partie précédente). Les annotations anaphoriques permettent d'identi-

TABLE 2.9 – Estimation de l’accord inter-annotateurs pour les unités lexicales et les tours de parole.

	A1	A2	<i>sm</i>	<i>pm</i>	$F_S$	$F_H$	$SER_S$	$SER_H$	Kappa
tours de parole	128	128	126	1	98.4	98.8	2.3	2.0	
entités personne	809	839	774	17	93.9	95.0	10.1	9.0	0.787
énumérations	102	100	68	16	67.3	75.2	49.5	41.6	0.642
entités temporelles	54	76	46	4	70.8	73.8	52.3	49.2	0.691
entités spatiales	85	86	63	8	73.7	78.4	43.3	38.6	0.716

fier les personnages rencontrés dans les contes. En accord avec les recommandations formulées par [Artstein and Poesio \(2008\)](#) et [Passonneau \(2004\)](#), l’accord sur les *chaines de coréférences* (définies dans la section 2.5.2) est mesuré à l’aide de l’alpha de Krippendorff ([Krippendorff, 1980](#)), défini par l’équation suivante :

$$\alpha = 1 - \frac{D_o}{D_e} \quad (2.8)$$

Avec  $D_o$  la mesure du désaccord observé, et  $D_e$  le désaccord attendu en cas d’annotation aléatoire, définis par :

$$D_o = \sum_{u=1}^N \frac{m_u}{n} D_u = \frac{1}{n} \sum_{u=1}^N \frac{1}{m_u - 1} \sum_{i=1, j=1}^m d(c_{iu}, c_{ju})^2 \quad (2.9)$$

$$D_e = \frac{1}{n(n-1)} \sum_{u=1, w=1}^N \sum_{i=1, j=1}^m d(c_{iu}, c_{ju})^2 \quad (2.10)$$

Avec  $N$  le nombre d’unités identifiées,  $m$  le nombre d’annotateurs,  $m_u$  le nombre d’annotations associées à l’unité  $u$  (permet de gérer les annotations incomplètes),  $c_{iu}$  l’annotation donnée par l’annotateur  $i$  à l’unité  $u$ , et  $d$  la mesure de dissimilarité retenue pour comparer deux annotations.



Les métriques  $d_p$ , Jaccard ( $d_j$ ) et MASI ( $d_m$ ) (cf. équations 2.11 à 2.13) sont utilisées pour estimer la distance entre deux ensembles d'unités. L'utilisation de l'alpha de Krippendorff nécessite qu'une étiquette soit assignée à chaque unité par un minimum de deux annotateurs, ce qui le rend particulièrement adapté aux tâches pour lesquelles les annotations sont disponibles en grand nombre. Dans le cadre de cette analyse qui mesure l'accord basé sur deux annotateurs, l'utilisation de l'alpha de Krippendorff nécessite d'ignorer les unités n'ayant été identifiées que par un seul des annotateurs (les entités nommées de personnes et les tours de parole).

$$d_p = \begin{cases} 0 & \text{si } A = B \\ 1 & \text{si } A \cap B = \emptyset \\ 1/3 & \text{si } A \subset B \text{ or } B \subset A \\ 2/3 & \text{autrement} \end{cases} \quad (2.11)$$

$$d_j = 1 - \frac{|A \cap B|}{|A \cup B|} \quad (2.12)$$

$$d_m = d_p * d_j \quad (2.13)$$

Le protocole utilisé par [Uzuner et al. \(2012\)](#) consiste à utiliser une adaptation de la F-Mesure pour estimer l'accord observé sur les annotations anaphoriques. Cette adaptation compare les paires de coréférences plutôt que les chaînes de coréférences. Elle est définie par l'équation 2.14.

$$F = \frac{2 * |Paires(A1) \cap Paires(A2)|}{|Paires(A1)| + |Paires(A2)|} \quad (2.14)$$

Deux variantes de cette mesure ont été utilisées :  $F_o$  qui correspond à la variante optimiste de la mesure, pour laquelle seules les unités identifiées par les annotateurs A1 et A2 sont prises en compte.  $F_p$ , la variante pessimiste de la mesure, qui prend en compte toute unité identifiée par au moins un des deux annotateurs et dont la mesure estime à la fois l'accord sur l'identification des unités et celui sur les annotations anaphoriques.

Les mesures d'accord associées aux annotations anaphoriques sont fournies dans la table 2.10. Quelle que soit la métrique utilisée avec l'alpha de Krippendorff ( $d_p$ ,  $d_j$  ou  $d_m$ ), le coefficient observé est supérieur à 0.8. Ceci permet de considérer que les annotations anaphoriques utilisées pour enrichir les entités de personne et les tours de parole sont fiables ([Krippendorff, 1980](#)). L'accord estimé pour les annotations anaphoriques des tours de parole est plus important que celui observé pour les entités de personne. Ces différences peuvent être partiellement expliquées par le fait que les unités correspondant aux tours de parole sont moins nombreuses, couvrent l'ensemble du texte et sont associées à des chaînes de coréférence plus courtes. Ces propriétés rendent la tâche d'annotation plus rapide, moins complexe et diminuent le risque d'erreur d'annotation. L'écart non négligeable observé entre les variantes pessimiste et optimiste de la F-Mesure est principalement lié aux désaccords portant sur l'identification des unités de personne présentées dans la table 2.9, plutôt que sur les annotations anaphoriques proprement dites. Malgré tout, les estimations

TABLE 2.10 – Accord inter-annotateurs estimé pour les annotations anaphoriques

	alpha $d_j$	alpha $d_p$	alpha $d_m$	$F_p$	$F_o$
Entités étendues personne	0.96	0.85	0.97	0.88	0.98
Tours de parole personnages	0.98	0.96	0.99	0.97	0.99

d'accord obtenues via la F-mesure restent bonnes et sont similaires à ce qui a pu être mesuré sur les entités personne lors de l'élaboration du corpus utilisé pour le concours i2b2/VA 2011 ([Uzuner et al., 2012](#)) (référencement de symptômes, personnes, procédures de test et traitements dans des documents médicaux).



## Chapitre 3

# Module d'extraction d'informations linguistique

### Sommaire

---

<b>3.1</b>	<b>Résumé du chapitre . . . . .</b>	<b>36</b>
<b>3.2</b>	<b>Annotation des unités lexicales . . . . .</b>	<b>36</b>
3.2.1	Signes de Ponctuation . . . . .	36
3.2.2	Parties du Discours . . . . .	37
3.2.3	Lemmatisation et Racinisation . . . . .	38
3.2.4	Synsets et Hyperonymes . . . . .	39
3.2.5	Structure Syntaxique . . . . .	39
3.2.6	Transcription phonétique . . . . .	41
<b>3.3</b>	<b>Module de détection automatique de tours de parole . . . . .</b>	<b>41</b>
3.3.1	Modèle de détection basé sur des automates simples . . . . .	42
3.3.2	Modèles basés sur des Champs Conditionnels Aléatoires Linéaires . . . . .	43
3.3.3	Évaluation . . . . .	47
3.3.4	Conclusion Partielle . . . . .	48
<b>3.4</b>	<b>Modèles de segmentation en épisodes . . . . .</b>	<b>50</b>
3.4.1	Modèles de segmentation en épisodes basés sur des automates . . . . .	51
3.4.2	Modèles des segmentation statistiques . . . . .	52
3.4.3	Conclusion partielle . . . . .	55
<b>3.5</b>	<b>Module d'extraction des actes de dialogue et de modes de communication . . . . .</b>	<b>56</b>
3.5.1	Modes de communication . . . . .	56
3.5.2	Actes de dialogue . . . . .	57
3.5.3	Conclusion partielle . . . . .	57

---

## 3.1 Résumé du chapitre

Ce chapitre décrit les différentes composantes du module d'analyse linguistique destiné à l'annotation automatique de texte de contes. Ce module prend en entrée un texte normalisé et produit un texte enrichi, dans un format XML correspondant à un enrichissement de la DTD utilisée pour représenter les contes. La première composante du module consiste à annoter les unités lexicales à l'aide de procédures externes, qui permettent d'extraire les parties du discours, les lemmes, les racines, les synsets, la structure syntaxique et la transcription phonémique des mots. La deuxième composante consiste à détecter les tours de parole (les citations directes) à l'aide de modèles basés sur des champs conditionnels aléatoires, qui sont entraînés et évalués sur le corpus texte décrit dans le chapitre précédent. Différents modèles, destinés à opérer une segmentation des contes en séquences d'épisodes, sont proposés puis évalués sur le corpus. Ces modèles sont basés sur l'utilisation des marques de paragraphe, des informations relatives aux tours de parole et des références aux personnages. La dernière composante du module linguistique vise à enrichir les phrases attribuées aux personnages à l'aide d'actes de dialogues et de modes de communication. Les modèles proposés sont basés sur des expressions régulières prenant en compte le contexte immédiat des tours de parole.

## 3.2 Annotation des unités lexicales

On désignera par *unités lexicales* les mots et signes de ponctuation identifiés au cours du processus de *tokenisation* décrit dans la section 2.3.2. Ce processus d'annotation consiste à associer à chaque unité lexicale des informations obtenues par des procédures externes. Les informations extraites sont mises au format XML à l'aide des balises **word** et **punct**, selon qu'il s'agit d'un mot ou d'un signe de ponctuation. Chaque information extraite est stockée dans un attribut de la balise.

### 3.2.1 Signes de Ponctuation

Les signes de ponctuation sont associés à un attribut **pctype** qui décrit le signe observé. Le rôle des guillemets est précisé, en utilisant les valeurs **lquote** ou **rquote**, selon qu'il s'agit d'un guillemet ouvrant ou fermant. Les marques de frontière de phrase et de paragraphe sont remplacées par les chaînes de caractères `\n` et `\n\n`. Ces chaînes sont considérées comme des signes de ponctuation et balisées avec des marqueurs `<punct pctype="carriage_return">`. Lorsque plusieurs signes de ponctuation identiques se suivent, ils sont regroupés au sein d'une seule balise portant l'attribut de ce signe de ponctuation, comme c'est le cas du texte balisé de l'exemple 9.

(9)

```
<punct ptype="exclamation">!/</punct> <punct ptype="exclamation">!/</punct> <punct
ptype="exclamation">!/</punct> <punct ptype="exclamation">!/</punct>
est transformé en
<punct ptype="exclamation">!!</punct>
```

### 3.2.2 Parties du Discours

Les *parties du discours* permettent de classer les mots rencontrés en neuf catégories principales : adjectif, adverbe, article, conjonction, interjection, nom, préposition, pronom et verbe. Selon la procédure d'analyse utilisée, des informations complémentaires sont utilisées pour décrire ces unités lexicales : mode, voix, personne, nombre et temps des verbes, le type de pronom (possessif, démonstratif, indéfini, personnel, relatif), etc. Ces parties du discours ont été obtenues grâce à deux procédures distinctes. Le premier système d'extraction utilisé est un outil interne du LIMSI. Ce système est basé sur des modèles de Markov cachés (HMM)(Rabiner and Juang, 1986) entraînés sur le corpus MULTITAG (Allauzen and Bonneau-Maynard, 2008). Le format d'encodage des informations extraites à l'aide de cette procédure est détaillé dans Rajman et al. (1997). La seconde procédure d'extraction des parties du discours est basée sur des arbres de décision probabilistes (Schmid, 1994). L'outil *TreeTagger*<sup>1</sup> est utilisé, en se servant des modèles entraînés sur la langue française par Stein and Schmid (1995).

La table 3.1 décrit l'accord observé entre les deux systèmes d'annotation des parties du discours, en se restreignant aux neuf catégories principales. Pour chaque catégorie morphosyntaxique, le nombre d'unités pour lesquelles les deux systèmes sont en accord est mentionné. Le nombre d'*alarmes* correspond aux unités d'une catégorie donnée détectées par un seul système. L'accord est estimé à l'aide de la F-Mesure, définie par l'équation 3.1.

$$F = \frac{2 * nb \text{ accords}}{2 * nb \text{ accords} + Alarmes_{Tree-Tagger} + Alarmes_{HMM}} \quad (3.1)$$

L'accord estimé sur l'étiquetage varie en fonction des différentes catégories morphosyntaxiques observées. Pour la plupart des catégories, il est supérieur à 90%, à l'exception des interjections (61.1%) et des adjectifs (81.3%). Ces différences soulignent la difficulté liée à l'obtention de parties du discours par le biais de procédures automatiques. Cette difficulté peut être partiellement expliquée par les spécificités du matériel utilisé, à savoir des contes pour enfants. La plupart des systèmes d'extraction d'information sont optimisés pour fonctionner sur des articles de journaux ou des textes médicaux (Maarouf and Villaneau, 2012; Goh et al., 2012) et non pas sur des textes de fiction. À défaut de pouvoir connaître le système d'étiquetage le plus adapté au cas du conte pour enfants, les informations obtenues peuvent être combinées pour accroître la robustesse des systèmes dont le fonctionnement nécessite l'utilisation de parties du discours. La pertinence des

---

1. disponible en ligne <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/french-tagset.html>

TABLE 3.1 – Accord observé pour l'étiquetage des parties du discours entre un système à base de HMM et un système basé sur TreeTagger

Catégorie morpho-syntaxique	Nb. accords	Alarmes TreeTagger	Alarmes HMM	Estimation d'accord (F-Mesure)
pronom	9183	1259	463	91.4%
verbe	12261	806	340	95.5%
nom	13059	1127	767	93.2%
adjectif	3055	417	986	81.3%
déterminant	8367	103	1472	91.4%
adverbe	4715	699	263	90.7%
conjonction	3521	202	445	91.6%
interjection	116	72	76	61.1%
préposition	7648	311	185	96.9%

étiquetages obtenus via ces deux systèmes combinés sera évaluée dans les sections suivantes en mesurant l'impact de leur utilisation sur les performances des systèmes qui les utilisent (pour la segmentation en tours de parole et en épisodes).

### 3.2.3 Lemmatisation et Racinisation

La *lemmatisation* consiste à déterminer la forme canonique d'un mot (par exemple : chevaux → cheval). Pour les verbes il s'agit de l'infinitif, pour les noms et adjectifs il s'agit de la forme masculin singulier. La *racinisation*, également appelée désuffixation (en anglais *stemming*) consiste à obtenir la forme tronquée d'un mot, en supprimant ces suffixes et flexions (ex : chevaux → cheva).

Bien que les buts de ces procédures soient proches, elles présentent différents avantages et inconvénients. Les procédures de lemmatisation très souvent basées sur des dictionnaires de formes fléchies, ont tendance à échouer à la moindre faute d'orthographe rencontrée. Elles nécessitent de désambiguïser le mot considéré, ce qui peut entraîner des erreurs de résolution. Les procédures de racinisation ne nécessitent pas de prendre en compte le contexte des mots rencontrés, et ont l'avantage d'être moins sensibles aux fautes d'orthographe. Elles gèrent relativement mal les différentes flexions pouvant être rencontrées sur les verbes irréguliers.

Les lemmes ont été obtenus en utilisant l'algorithme TreeTagger (Schmid, 1994), utilisé avec le fichier de paramètres pour la langue française réalisé par Achim Stein (Stein and Schmid, 1995). Les racines des mots sont obtenues en utilisant les règles exprimées en langage *Snowball* (Porter, 2001) fournies dans NLTK (Loper and Bird, 2002). La table 3.2 recense quelques exemples illustrant les cas où l'usage de ces procédures peut être problématique.

TABLE 3.2 – Comparaison des racines obtenues avec NLTK et des lemmes obtenus avec TreeTagger

Phrase initiale	Racinisation	Lemmatisation
je suis en retard	je suis en retard	je suivre être en retard
tu es en retard	tu e en retard	tu être en retard
il est en retard	il est en retard	il être en retard
nous sommes en retard	nous somm en retard	nous sommer être en retard
vous êtes en retard	vous ête en retard	vous être en retard
ils sont en retard	il sont en retard	ils être en retard
un instituteur	un instituteur	un instituteur
une institutrice	une institutric	un instituteur

### 3.2.4 Synsets et Hyperonymes

Les *synsets* (de l'anglais : *synonym set*, ensemble de synonymes) sont des ensembles de mots interchangeables dans un contexte donné (exemple : manger, déjeuner, se nourrir, grignoter, ...). Les mots polysémiques (qui ont plusieurs sens) appartiennent à plusieurs synsets. Les *hyperonymes* sont des catégories dont le sens inclut celui d'un autre mot (fruit est un hyperonyme de kiwi, pomme, orange).

L'annotation des synsets et hyperonymes a été réalisée en utilisant la base de données WOLF (Sagot et al., 2008), qui est une traduction en Français du projet anglophone Wordnet (Miller, 1995). Chaque mot de la base est associé à une liste de synsets correspondant à ses différentes significations. La probabilité d'appartenance d'un mot à un synset est également mentionné. Les synsets y sont organisés sous forme d'*ontologie* : c'est-à-dire sous forme d'arbre défini tel que le père d'un noeud soit son hyperonyme.

L'étiquetage des synsets et des hyperonymes a consisté à faire correspondre les mots aux synsets pour lesquels ils ont la plus grande probabilité d'appartenance (sans prise en compte du contexte d'apparition). Chaque mot a été associé à l'identifiant numérique de son synset, ainsi qu'aux identifiants de ses hyperonymes pour des profondeurs fixées.

### 3.2.5 Structure Syntaxique

La structure syntaxique des textes est représentée sous forme d'arbre syntaxique abstrait, comme illustré par la figure 3.1. L'arbre est obtenu à partir de règles opérant sur les parties du discours obtenues à l'aide du système décrit précédemment (Allauzen and Bonneau-Maynard, 2008). Les règles de construction de l'arbre sont définies pour être utilisées avec le moteur d'expressions régulières WMATCH (Galibert, 2009; Rosset et al., 2009). Ce moteur permet de générer des structures arborescentes à partir d'expressions régulières définies sur des mots, ou d'autres structures arborescentes. La structure arborescente est ensuite convertie de manière à pouvoir être représentée à l'aide d'un attribut multidimensionnel par mot, comme illustré par la figure 3.2.



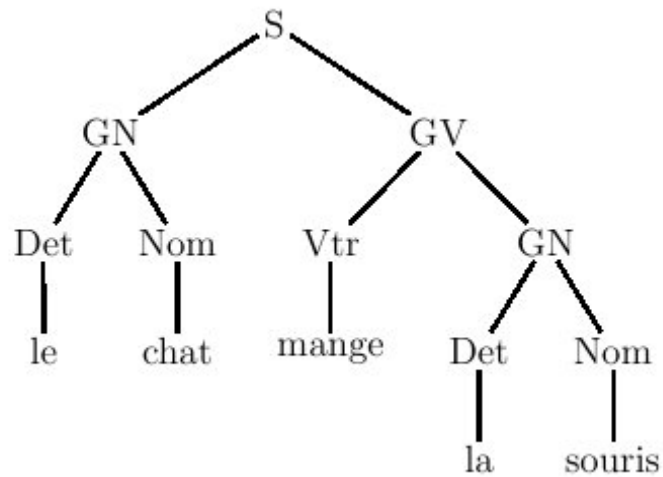


FIGURE 3.1 – Arbre syntaxique abstrait correspondant à la phrase “le chat mange la souris”. P est une phrase, GN un groupe nominal, GV un groupe verbal

S	GN	Det	le	le
S	GN	Nom	chat	chat
S	GV	Vtr	mange	mange
S	GV	GN	Det	la
S	GV	GN	Nom	souris

FIGURE 3.2 – Arbre syntaxique abstrait équivalent à la figure 3.1 représenté à l’aide d’un attribut multidimensionnel par mot

#### 3.2.6 Transcription phonétique

Le système de synthèse de la parole utilisé dans le cadre du projet GV-LEx procède à une conversion graphème-phonème permettant de sélectionner les unités de parole les plus adaptées à un contexte donné. Cependant, du fait de la structure des traitements automatique mis en œuvre dans le cadre du projet, il n'est pas possible d'accéder à la transcription phonétique utilisée par ce système - qui intervient sur le fichier XML créé par ces prétraitements. Cette transcription est nécessaire pour pouvoir utiliser des règles de contrôle prosodiques liées à la taille du contexte, exprimé en nombre de phonèmes. Le système libre à base de règles LIA\_PHON (Béchet, 2001) a donc été utilisé à cette fin.

### 3.3 Module de détection automatique de tours de parole

Cette section décrit les différents prototypes d'annotation automatique des tours de parole réalisés dans le cadre de cette thèse. Un tour de parole se définit comme une zone de texte contigüe qui correspond à une citation directe attribuée à un unique personnage (cf. section 2.4.3). La taille de cette zone est variable : il peut s'agir de quelques mots ou de quelques phrases. Une citation directe peut correspondre à ce que dit le personnage, comme à ce qu'il pense (cf. exemple 10). Une même phrase peut contenir plusieurs tours de parole (cf. exemple 11), entrecoupés par des *incises*, permettant d'identifier le personnage cité, ou la manière dont l'information citée a été formulée (pensée, hurlement, soupir...). Même si ce n'est pas systématique, les premières phrases des citations directes sont souvent indiquées par des signes de ponctuation spécifiques, tels que des tirets ou des guillemets (l'analyse du corpus a montré que plus de 2/3 des citations directes possédaient une telle ponctuation). Les systèmes de détection doivent être à même de déterminer si les phrases suivantes restent attribuées au personnage de la citation initiale ou si elles doivent être attribuées au narrateur (cf. exemple 12).

(10) alors il pensa :

- “si un être rencontre son ennemi à l'endroit où il va être sacrifié, il ne peut que se battre pour se défendre et ne pas abandonner.”

(11) - tu verras ma brave petite fille dit la sorcière en enveloppant Ludivine dans une cape, tu seras bien avec moi.

(12) La fée continua : - Je t'ordonne d'achever ce que tu as de plus cher aujourd'hui : ton cheval!

Bouleversé, le prince lui répondit :

- Mon cheval!

Jamais!

Les citations directes observées à l'oral sont généralement liées à un certain nombre de propriétés prosodiques (Holt, 1996). Ces effets se traduisent généralement par une hauteur plus élevée et un volume plus important. Dans le cas de la lecture de contes, il peut aussi s'agir d'effets destinés à

imiter un personnage type (Přibil and Přibilová, 2008). Un certain nombre d'approches visant à améliorer l'expressivité des systèmes de synthèse ont consisté à concevoir des systèmes à base de règles capables d'identifier les citations directes (cf. Zhang et al. (2003) pour l'Anglais et Mamede and Chaleira (2004) pour le Portugais), pour y associer manuellement dans un second temps des voix de synthèse correspondant aux personnages. Les annotations relatives aux tours de parole des personnages couvrent une partie importante du corpus texte GV-LEx (30.1% des mots) et sont associées à un très fort accord inter-annotateur. Ces propriétés permettent d'utiliser le corpus pour entraîner et évaluer des modèles d'annotation automatique.

### 3.3.1 Modèle de détection basé sur des automates simples

Deux automates basés sur des règles simples ont été conçus pour détecter des tours de parole. Ces automates servent de base à la description des performances obtenues à l'aide de modélisations plus complexes.

Les deux modèles proposés détectent les tours de parole dans des zones de texte associées à des marqueurs de ponctuation tels que les guillemets et les tirets. Ces zones sont définies, soit comme des phrases (obtenues lors du processus de tokenisation) commençant par un tiret, soit comme des zones de texte comprises entre deux guillemets. Ces définitions ont été étendues afin de prendre en compte différents cas de figure rencontrés dans le corpus : la présence de guillemets suivis ou précédés par un tiret ou des guillemets fermants suivis par des signes de ponctuation terminaux (indiquant la fin d'une phrase). Un nouveau tour de parole de personnage est prédit au commencement de chacune de ces zones. Les changements de personnage au sein d'une zone entourée par des guillemets sont liés à la présence de tirets, et associés à la création de nouveaux tours de parole.

L'automate le plus simple, appelé *TPAutoPunct*, n'utilise que les signes de ponctuation pour déterminer la présence de tours de parole. Un deuxième automate, *TPAutoPOS*, utilise les annotations de parties du discours obtenues à l'aide du modèle à base de HMM proposé par Allauzen and Bonneau-Maynard (2008) et présenté dans la section 3.2.2. Les parties du discours sont utilisées pour détecter les *incises*, qui sont associées à des changements de tour de parole au sein d'une même phrase normalisée. Les règles de détection des incises consistent à repérer les signes de ponctuation suivis, soit par des verbes, soit par des pronoms personnels datifs (lui) ou réflexifs (se) à la troisième personne et suivis par des verbes. La détection d'une incise a pour effet de fermer le tour de parole du personnage, et de créer un tour de parole du narrateur. Un tour de parole de personnage supplémentaire est créé si un signe de ponctuation est détecté après l'incise. Le comportement des deux automates est illustré par l'exemple 13.

(13) Annotation en tours de parole obtenue avec l'automate *TPAutoPunct*

<spkr>-"est-ce que je suis ici venue, lui dit cette brutale orgueilleuse, pour vous donner à boire?"</spkr>

Annotation en tours de parole obtenue avec l'automate *TPAutoPOS*

<spkr>-"est-ce que je suis ici venue,</spkr> lui dit cette brutale orgueilleuse, <spkr>pour vous donner à boire?"</spkr>

#### 3.3.2 Modèles basés sur des Champs Conditionnels Aléatoires Linéaires

Les *champs conditionnels aléatoires* (Lafferty et al., 2001) sont des modèles probabilistes discriminants, adaptés aux données structurées sous forme de graphe. L'efficacité de leur formulation linéaire, restreignant leur utilisation aux séquences, a été attestée sur un grand nombre de tâches ayant trait au traitement automatique des langues : extraction des parties du discours (Lafferty et al., 2001; PVS and Karthik, 2007; Gimpel et al., 2011), ou encore détection d'entités nommées (McCallum, 2002; Settles, 2004). Cette partie décrit l'utilisation de ces modèles pour la tâche de segmentation des contes en tours de parole. Dans la suite de cette partie, les champs conditionnels aléatoires seront désignés à l'aide de leur acronyme anglais : **CRF** (Conditional Random Fields).

#### Champs Conditionnels Aléatoires Linéaires

Un CRF linéaire consiste à prédire la probabilité d'une séquence d'annotation  $y = (y_1, \dots, y_T)$  pour une séquence d'observations donnée  $x = (x_1, \dots, x_T)$ , tel que :

$$p(y|x) = \frac{1}{Z(x)} \exp\left(\sum_{t=1}^T \sum_k \lambda_k f_k(x, y, t)\right) \quad (3.2)$$

Les fonctions  $f_k$ , appelées *features*, sont des fonctions à valeur réelle. Il s'agit généralement de fonctions binaires qui retournent 1 si un phénomène est observé, 0 sinon. Elles sont systématiquement définies par l'utilisateur d'un modèle CRF pour intégrer ses connaissances sur les phénomènes à modéliser. Les coefficients  $\lambda_k$  sont assimilables à des *poids* permettant d'accorder plus ou moins d'importance aux différentes features. Il s'agit de paramètres généralement appris via une procédure d'entraînement sur des données annotées.  $Z(x)$  est une fonction de normalisation définie telle que :

$$Z(x) = \sum_y \exp\left(\sum_{t=1}^T \sum_k \lambda_k f_k(x, y, t)\right) \quad (3.3)$$

Lorsque les poids d'un modèle CRF sont fixés, celui-ci est utilisé pour prédire les probabilités associées à l'ensemble des séquences d'annotations possibles. Le choix de la séquence d'annotation optimale est généralement réalisé via un algorithme de programmation dynamique, tel que l'algorithme de Viterbi (Viterbi, 1967).

## Formalisme d'annotation des tours de parole

Les tours de parole sont des segments contigus. Leur utilisation avec des CRF nécessite de définir un formalisme de représentation de ces segments, basé sur un nombre fini de symboles. Chaque observation (en l'occurrence : les mots et les signes de ponctuation des textes) est associée à un symbole  $y_i$  (équation 3.2). Pour cela, un schéma d'étiquetage inspiré par les représentations BIO<sup>2</sup>, composé de quatre étiquettes, (Ramshaw and Marcus, 1995) est défini. Les étiquettes préfixées par *BEGIN* indiquent un nouveau tour de parole. La distinction entre le narrateur (*NARR*) et les personnages (*SPKR*) est précisée par le suffixe de l'étiquette. L'ensemble des étiquettes définies est listé ci-dessous. L'exemple 14 illustre l'utilisation de ce formalisme sur une phrase.

**BEGIN\_NARR** : Début d'un segment tour de parole du narrateur (début d'un texte, ou fin d'un tour de parole de personnage)

**INSIDE\_NARR** : Intérieur d'un segment tour de parole du narrateur

**BEGIN\_SPKR** : Début d'un segment tour de parole de personnage (après un tour de parole du narrateur, ou d'un personnage différent)

**INSIDE\_SPKR** : Intérieur d'un segment tour de parole narrateur

(14)

j'	BEGIN_SPKR
ai	INSIDE_SPKR
faim	INSIDE_SPKR
,	INSIDE_SPKR
dit	BEGIN_NARR
la	INSIDE_NARR
souris	INSIDE_NARR
,	INSIDE_NARR
et	BEGIN_SPKR
j'	INSIDE_SPKR
ai	INSIDE_SPKR
soif	INSIDE_SPKR
!	INSIDE_SPKR

## Définition des fonctions features utilisées pour la tâche de segmentation en tours de parole

Dans le cadre de cette étude, les fonctions features utilisées sont des fonctions booléennes qui retournent 1 si une propriété est observée, 0 sinon. Les observations  $x_i$  peuvent être multi-dimensionnelles : les différentes dimensions d'une observation peuvent correspondre au token brut observé (mot ou signe de ponctuation), ou à un descripteur quelconque tel que défini dans la section 3.3.2 (partie du discours, lemme, racine...). Quelques exemples de fonction features sont décrits ci-dessous :

---

2. de l'anglais *Begin Inside Outside*

- $f_1(x, y, t) = 1$  si  $x_t[token] = j'$  et  $y_t = BEGIN\_SPKR$
- $f_2(x, y, t) = 1$  si  $y_t = BEGIN\_SPKR$  et  $y_{t-1} = INSIDE\_NARR$  et  $x_t[token] = j'$
- $f_3(x, y, t) = 1$  si  $y_t = BEGIN\_SPKR$  et  $y_{t-1} = INSIDE\_NARR$   
et  $x_t[partiedudiscours] = \textit{pronom personnel première personne}$   
et  $x_t[partiedudiscours] = \textit{verbe présent première personne}$

La fonction  $f_1$  s'active si le token observé est "j'" et si l'annotation associée est un début de tour de parole de personnage. La fonction  $f_2$  prend en considération la valeur de l'annotation courante ainsi que celle de l'annotation précédente. On parle alors de bigramme d'annotation. Elle s'active en cas de transition narrateur vers personnage, si le token courant est égal à "j'". La fonction  $f_3$  est également définie par rapport à un bigramme d'annotation. Elle s'active si le bigramme correspondant aux parties du discours de l'observation courante et suivante est un pronom personnel première personne suivi d'un verbe conjugué à la première personne du présent.

### Patron de définition de fonction features

Le nombre de fonctions features définies dans le cadre d'une tâche de traitement des langages naturels peut être très grand. Il est courant de définir une fonction feature pour chaque observation, ou bigramme d'observations, rencontrés dans les données d'entraînement : qu'il s'agisse d'unités lexicales (tokens) ou de descripteurs (parties du discours, lemmes...) En conséquence, les utilisateurs de modèles CRF définissent généralement des *patrons* permettant de générer des fonctions features à partir des données d'entraînement. L'exemple 15 décrit un *patron* d'extraction de fonction features, défini pour des observations de dimension 3, telle que la première dimension corresponde au token brut, la deuxième à la partie du discours, et la troisième au lemme. Chaque ligne représente une règle d'extraction de feature. Le premier symbole peut prendre les valeurs

**U** : si la feature s'active pour la valeur de l'annotation courante ( $y_t$ )

**B** : si la feature s'active pour la valeur du bigramme correspondant à l'annotation courante et l'annotation précédente ( $y_{t-1}/y_t$ )

**\*** si l'on souhaite créer une feature prenant en compte la valeur de l'annotation courante (U) et une autre feature qui prenne en compte la valeur du bigramme (B).

Les observations  $x[i, j]$  sont indicées par leur position relative à l'observation courante (i) et la dimension de l'observation considérée (j).

(15)

U :  $x[0,0]$  % token courant, annotation couranteB :  $x[0,0]$  % token courant, bigramme annotation précédente/courante\* :  $x[-1,0]/x[0,0]$  % bigramme token précédent/token courant, annotation courante + bigramme annotation précédente/courante\* :  $x[0,0]/x[1,0]$  % bigramme token courant/token suivant\* :  $x[-1,0]/x[0,0]/x[1,0]$  % trigramme token précédent/courant/suivant\* :  $x[-1,1]/x[0,1]$  % bigramme partie du discours précédente/courante\* :  $x[0,1]/x[0,2]$  % bigramme partie du discours/lemme courants**Entraînement de modèles CRF sur le corpus GV-LEx**

L'espace des paramètres d'entraînement des modèles CRF est exploré de manière exhaustive afin de trouver le modèle le plus adapté à la tâche de segmentation en tours de parole.

Le premier axe d'exploration consiste à déterminer une combinaison optimale de descripteurs associés aux unités lexicales. En d'autres termes, il s'agit de déterminer le nombre de dimension des observations  $x_i$ , et le type d'information stocké dans chaque dimension. Un total de treize descripteurs distincts sont considérés : unité lexicale brute (mot ou signe de ponctuation), présence de lettre majuscule (descripteur booléen), distinction entre mots et signes de ponctuation (booléen), racines et suffixes, lemmes, structure syntaxique, parties du discours (détaillés dans la section ). En fonction du niveau de détail des parties du discours considérées, celles-ci peuvent être scindées en plusieurs descripteurs distincts, afin d'améliorer les capacités de généralisation des modèles. Les parties du discours obtenues à l'aide de l'algorithme TreeTagger (Stein and Schmid, 1995) sont les moins détaillées (on observe 29 étiquettes différentes). Elles contiennent la catégorie syntaxique principale (nom, verbe) pouvant être concaténée avec un symbole permettant de détailler l'entité lexicale (temps de conjugaison, type de pronom). Les parties du discours obtenues à l'aide du système par HMM (Allauzen and Bonneau-Maynard, 2008) sont davantage détaillées (232 étiquettes différentes observées). Chaque étiquette obtenue via cette procédure est utilisée pour générer quatre descripteurs supplémentaires : la catégorie principale (nom, verbe, ...), la personne (pour les verbes, pronoms, déterminants), l'étiquette simplifiée (c'est-à-dire ne contenant plus d'informations relatives au genre, au nombre ou à la personne) et la séparation de l'étiquette en sept descripteurs atomiques utilisés de manière combinée (cf. Rajman et al., 1997). Un total de 24 combinaisons de descripteurs est ainsi évalué sur le corpus.

Le deuxième axe d'exploration de l'espace des paramètres consiste à déterminer l'impact de la taille des différentes fenêtres d'observation. Les observations sont systématiquement associées à l'annotation courante, ainsi qu'au bigramme contenant l'annotation précédente et l'annotation courante. Pour chaque dimension, l'ensemble des unigrammes d'observations est pris en compte sur des intervalles contextuels dont la taille varie entre  $(-2, +2)$  à  $(-9, +9)$ . L'ensemble des observations regroupées en bigrammes sont prises en compte sur des fenêtres dont la taille varie entre

(0, 0) à (−3, +3). Des expériences supplémentaires non concluantes ont été réalisées en regroupant les observations en trigrammes. Elles ne seront pas détaillées dans la suite de ces travaux.

La suite d’outils `Wapiti 1.1.2` (Lavergne et al., 2010) est utilisée pour entraîner des modèles CRF sur le corpus texte. L’algorithme d’entraînement sélectionné pour entraîner les modèles est L-BFGS (Liu and Nocedal, 1989). Pour chacune de ces combinaisons, trois valeurs de critères d’arrêt et trois valeurs de pénalité elastic-net L1 (Zou and Hastie, 2005) sont évaluées. L’ensemble des combinaisons de ces différents paramètres (ensemble de descripteurs, taille contextuelle des unigrammes, taille contextuelle des bigrammes, critères d’arrêt et pénalité elastic-net L1) représente un total de 6912 modèles. L’entraînement et l’évaluation de ces modèles est réalisé sur un cluster de calcul `Sun Grid Engine` (Etsion and Tsafir, 2005).

### 3.3.3 Évaluation

Le corpus texte GV-LEx a été utilisé pour l’entraînement et l’évaluation des différents modèles présentés. L’ensemble des annotations disponibles est utilisé, à l’exception des annotations doublons (contes annotés par plusieurs annotateurs). Les critères de sélection des annotations doublons sont identiques à ceux décrits dans la section 2.5. Les modèles nécessitant des données d’apprentissage sont évalués en utilisant un processus de validation croisée à 10 échantillons<sup>3</sup> (Kohavi et al., 1995).

Les métriques utilisées pour évaluer les performances des modèles sont la F-Mesure stricte (équation 2.4), ainsi que le Slot Error Rate (SER) (Makhoul et al., 1999) défini par l’équation 3.4.

$$SER(ref, hyp) = 100 \frac{D + I + FB}{|ref|} \quad (3.4)$$

Où  $|ref|$  représente le nombre d’unités (tours de parole personnage) trouvés dans l’annotation de référence,  $D$  la quantité de suppressions (faux négatifs : tours de parole personnage non détectés par le modèle),  $I$  la quantité d’insertions (faux positifs : unités prédites par le modèle, mais absentes dans la référence) et  $FB$  le nombre de paires d’unités se chevauchant et ayant des frontières différentes. Les meilleurs modèles doivent combiner une F-Mesure élevée et un SER faible.

Les performances obtenues par les différents modèles sont détaillées dans la table 3.3. Les automates *TPAutoPunct* et *TPAutoPOS* permettent de détecter respectivement 41% et 54% des tours de parole du corpus. Les règles basées sur l’utilisation des parties du discours utilisées dans *TPAutoPunct* permettent de réduire les erreurs liées à la non-détection des incises. Une limitation prévisible de ces modèles se trouve dans leur incapacité à détecter les tours de parole des personnages hors des zones associées à des tirets ou des guillemets, qui couvrent 12% du corpus et presque un tiers des tours de parole (cf. section 2.5). Une autre limitation majeure de ces modèles vient de leur incapacité à déterminer si les phrases qui suivent la première phrase d’un personnage font partie ou non du même tour de parole, si elles ne sont pas délimitées par des guillemets.

---

3. 10-fold cross validation



Les performances des modèles s'appuyant sur des CRF sont regroupées par type de descripteur utilisé :

**O** : Descripteurs orthographiques (chaîne de caractères brute, majuscules et distinction mot/-  
signe de ponctuation)

**HPOS** : Parties du discours obtenues avec le système basé sur des HMM

**TPOS** : Parties du discours obtenues avec le système basé sur TreeTagger

**STM** : Racinisation et/ou suffixes

**SP** : Structure syntaxique

Seules les performances obtenues avec le meilleur modèle de chaque catégorie (taille contextuelle et paramètres d'entraînement fournissant la meilleure combinaison) sont retranscrites, afin de décrire essentiellement l'impact des différents descripteurs. Les meilleures performances proviennent du modèle qui combine les parties du discours obtenues avec les deux systèmes (cf. section 3.2.2). L'ajout d'informations relatives à la racinisation ou à la structure syntaxique ne permet pas d'obtenir de meilleurs résultats. Suite à des analyses complémentaires, il s'avère que ces informations permettent d'améliorer les performances des modèles lorsqu'elles utilisées avec des faibles tailles contextuelles, mais se révèlent inefficaces combinées avec des grandes tailles contextuelles. La figure 3.3 illustre les meilleures performances obtenues par les modèles, tous ensembles de descripteurs confondus, pour des tailles contextuelles fixées. Les tailles contextuelles optimales sont observées pour des contextes d'unigrammes compris entre  $(-6, +6)$  et  $(-9, +9)$ , combinés avec des contextes de bigrammes compris entre  $(-1, +1)$  et  $(-2, +2)$ . Pour toutes les tailles contextuelles prises en compte dans cette évaluation, les modélisations à base de CRF permettent d'obtenir des résultats nettement supérieurs à ceux obtenus par les automates de base. La modélisation à base de CRF la plus simple, n'utilisant que les descripteurs orthographiques et la plus petite taille contextuelle prise en compte  $(-2, +2)$  permet d'obtenir une F-Mesure de 81,3 associée à un SER de 24,9. Le meilleur modèle obtenu utilise des unigrammes de taille contextuelle égale à  $(-9, +9)$ , associés à des bigrammes de taille contextuelle égale à  $(-2, 2)$ . Il permet d'obtenir une F-mesure de 88,8 associée à un SER de 15.1. Il permet de détecter correctement 86% des tours de parole des personnages. L'analyse des erreurs de prédiction montre que 9% des erreurs sont dues à des insertions de tour de parole, 20% à des suppressions et 71% à des erreurs de frontière.

### 3.3.4 Conclusion Partielle

La quantité suffisante d'annotations de citations directes dans le corpus GV-LEx a permis de proposer, d'entraîner et d'évaluer des systèmes de détection des citations directes. Les modélisations à base de CRF ont permis d'obtenir des résultats encourageants. Leur utilisation peut être intéressante hors du cadre de la synthèse de parole, par exemple pour la conception de système de question réponse, ou de procédures d'analyse du discours, sous réserve que la robustesse des modèles proposés soit évaluée sur des ressources qui ne soient pas de textes de fiction.

TABLE 3.3 – Résultats obtenus par les automates et les modélisations à base de CRF pour la tâche de segmentation en tours de parole.

Modèle	F-Mesure	SER
TPAutoPunct	40.2	80.0
TPAutoPOS	50.9	70.4
O	83.7	21.4
O + HPOS	86.8	17.8
O + TPOS	88.2	16.0
O + HPOS + TPOS	88.9	15.1
O + HPOS et/ou TPOS + SP et/ou STM	88.9	15.3

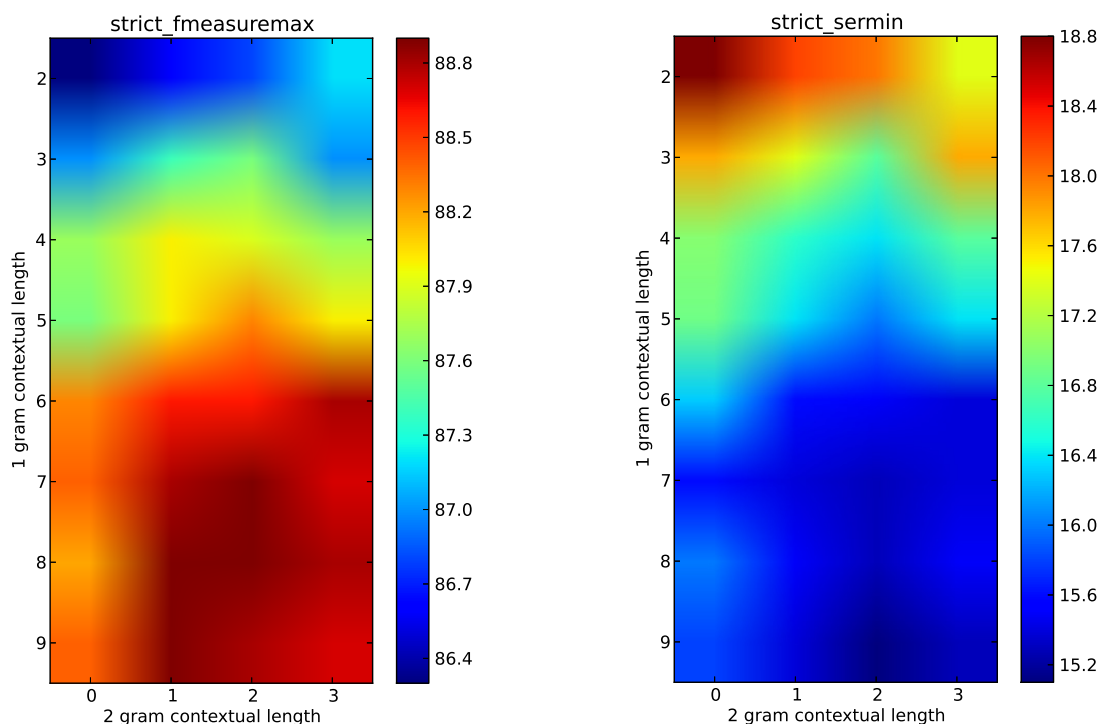


FIGURE 3.3 – Variation des performances obtenues pour la tâche de segmentation en tours de parole par les modélisations basées sur des CRF, pour des tailles contextuelles d’unigrammes et de bigrammes fixées

Dans le cadre de tests préliminaires, il a été observé que l'utilisation des *synsets* et des *lemmes* (sections 3.2.4 et 3.2.3) a détérioré les performances des modèles. La procédure d'étiquetage des *synsets* peut en partie expliquer ce phénomène : un même mot peut en effet correspondre à plusieurs *synsets*. Le choix du *synset* optimal nécessite une analyse du contexte dans lequel le mot est employé, qui n'a pas été réalisée. Par soucis de concision, les expériences utilisant les lemmes et *synsets* ne sont pas mentionnées plus avant dans ce manuscrit.

Plusieurs pistes d'amélioration des modèles proposés sont envisageables. Les meilleurs modèles obtenus combinent des annotations des parties du discours obtenues via deux procédures automatiques différentes (Allauzen and Bonneau-Maynard, 2008; Stein and Schmid, 1995). Il semble donc vraisemblable que l'utilisation de procédures d'étiquetage des parties du discours supplémentaires, telle que l'analyseur de Stanford (Toutanova et al., 2003), puissent améliorer les performances des modèles. D'autres approches consisteraient à utiliser des dictionnaires de verbes généralement associés à des citations directes, ou encore à utiliser les prédictions de systèmes à base de règles (par exemple Weiser and Watrin, 2012) comme des descripteurs supplémentaires qui seront utilisés par les modélisations basées sur des CRF.

### 3.4 Modèles de segmentation en épisodes

Cette partie décrit différents modèles qui permettent de segmenter les contes en séquences d'épisodes dont l'efficacité est évaluée sur le corpus texte GV-LEx. Comme cela déjà été mentionné, les titres des contes correspondent systématiquement à la première ligne des documents, séparée du reste du texte par des marqueurs de paragraphes. Leur identification est triviale, et pour cette raison, ils sont ignorés dans les expériences ci-dessous. Les modèles proposés utilisent différentes informations : les marques de paragraphes, les annotations en tours de parole et les entités personnage enrichies avec des annotations anaphoriques. Les marques de paragraphes sont obtenues de manière automatique à l'issue du processus de tokenisation. Les annotations en tour de parole sont obtenues automatiquement en utilisant le meilleur modèle à base de CRF décrit dans la section précédente. Les annotations utilisées correspondent aux annotations générées lors du processus de validation croisée ; en d'autres termes les annotations en tour de parole qui correspondent à un conte donné sont obtenues à l'aide d'un modèle qui n'a pas utilisé le conte dans sa base d'apprentissage. Les entités personnage sont obtenues à partir des annotations manuelles. Leur utilisation ne permet pas d'obtenir des prototypes utilisables en l'état, mais leur évaluation permet de suggérer des pistes d'amélioration des modèles de segmentation.

La qualité des segmentations obtenues à l'aide des différents modèles est évaluée à l'aide de plusieurs métriques. Le Ratio de Frontières Épisodiques (**RFE**) défini comme le nombre de frontières observées dans une segmentation prédite, divisé par le nombre de frontières observées dans la segmentation de référence. Un RFE supérieur à 1 signifie que la segmentation obtenue à l'aide d'un modèle contient plus de frontières que la segmentation de référence. La précision, définie comme le ratio de frontières prédites qui sont observables dans la segmentation de référence. Le rappel,

TABLE 3.4 – Évaluation des modèles de segmentation en épisodes basés sur des automates

Segmenteur	RFE	Précision	Rappel	F-Mesure	WD	Pk
SegVide	0.00	inf	0.00	0.00	<b>0.43</b>	0.43
SegSyst	7.65	0.13	<b>1.00</b>	0.23	1.00	0.57
SegAléatoire	<b>1.00</b>	0.12	0.12	0.12	0.56	0.51
Fp	2.27	0.31	<b>0.71</b>	0.43	0.57	0.40
FpTpManu	1.45	<b>0.42</b>	0.61	<b>0.49</b>	0.47	<b>0.36</b>
FpTpAuto	1.46	<b>0.41</b>	0.60	<b>0.49</b>	0.48	<b>0.36</b>

défini comme le ratio de frontières contenues dans la segmentation de référence qui sont observées dans la segmentation prédite par un modèle. La F-mesure, définie comme la moyenne harmonique de la précision et du rappel. Les mesures WindowDiff et Pk (mesures détaillées à la section 3.4.3), utilisées avec une taille de fenêtre  $k = 4$ . À l'exception de la RFE, toutes ces métriques fournissent une valeur réelle comprise entre 0 et 1. Les meilleurs modèles sont ceux permettant d'optimiser l'ensemble des métriques, c'est-à-dire à combiner les plus grands rappels, précision et F-Mesure, avec les plus basses valeurs de WindowDiff et Pk et un RFE le plus proche de 1.

### 3.4.1 Modèles de segmentation en épisodes basés sur des automates

Les frontières des épisodes sont généralement induites par des changements de sujet, l'apparition d'un nouveau personnage, un déplacement spatial, ou un changement temporel dans le cours du récit. Ces changements sont souvent indiqués typographiquement grâce à des stratégies d'indentation, se traduisant par des marques de paragraphe ou des retours à la ligne dans le texte normalisé. Cependant, il arrive que les marques de paragraphe soient insérées par l'auteur dans le seul but d'aérer l'apparence du texte et ainsi de faciliter la lecture (Hearst, 1997).

Dans l'expérience ci-dessous, les relations entre marqueurs de paragraphes et frontières d'épisodes sont explorées à l'aide d'automates. Un modèle naïf consiste à associer chaque marque de paragraphe à une frontière d'épisode (**Fp**). Dans le cadre d'analyses informelles du corpus, il a été observé que les frontières de paragraphe suivies par des citations directes de personnage étaient rarement associées à des frontières d'épisode. Cette observation est formalisée à l'aide d'automates qui n'associent que les marques de paragraphes suivies par un tour de parole du narrateur aux frontières d'épisodes : **FpTpManu** et **FpTpAuto**, modèles qui utilisent respectivement les tours de parole obtenues à partir des annotations manuelles et ceux obtenus à l'aide d'une procédure automatique basée sur des CRF. Afin de faciliter la description de la qualité des segmentations obtenues et de réduire les biais spécifiques aux différentes métriques d'évaluation considérées, trois segmentations dégénérées sont également décrites. Une segmentation aléatoire (**SegAléatoire**), qui contient le même nombre de frontières que la segmentation de référence. Une segmentation vide (**SegVide**), ne contenant aucune frontière. Une segmentation telle que chaque frontière de phrase soit systématiquement associée à une frontière d'épisode (**SegSyst**).

La table 3.4 décrit l'estimation de la qualité des différentes segmentations obtenues à l'aide de

l'ensemble des métriques décrites ci-dessus. La valeur de rappel associée à l'automate n'utilisant que les marques de paragraphe (Fp) montre que 71% des frontières des épisodes sont associées à des marques de paragraphe. La conséquence directe de cette observation est qu'un modèle de segmentation en épisodes dont les prédictions seraient restreintes aux frontières de paragraphe n'est pas à même de détecter 29% des frontières d'épisode. La valeur de précision associée à l'automate Fp montre que plus des deux tiers des marques de paragraphe ne sont pas associées à des frontières d'épisodes. La combinaison des marques de paragraphe et des annotations relatives aux tours de parole permet d'améliorer la qualité des segmentations estimées par l'ensemble des métriques, à l'exception du rappel. Les résultats obtenus en utilisant les tours de parole provenant des annotations manuelles ou des procédures automatiques sont très similaires et permettent de considérer que les procédures de segmentation automatique en tours de parole sont suffisamment robustes pour être utilisées par défaut dans la suite des essais de segmentation en épisode. Les segmentations basées sur les marques de paragraphes et/ou sur les informations liées aux tours de parole sont jugées de meilleure qualité que les segmentations dégénérées par les métriques F-Mesure et Pk. Leur important taux de fausses alarmes provoque un score de WindowDiff de moindre qualité que celui obtenu en considérant la segmentation vide.

### 3.4.2 Modèles des segmentation statistiques

Malgré leurs performances raisonnables, les automates proposés dans la partie précédente présentent un certain nombre de limitations. Leur utilisation est restreinte aux textes préalablement segmentés en paragraphes. Cette limitation n'est pas forcément gênante si l'on se situe dans des scénarios pour lesquels cette information est disponible, mais elle exclue 29% des frontières d'épisodes annotées de l'ensemble des prédictions possibles. Dans le cas d'un automate utilisant conjointement les marques de paragraphe et les tours de parole, c'est alors 39% des frontières annotées manuellement qui ne sont pas détectées. De plus, ces modèles présentent une tendance à identifier un plus grand nombre de frontières d'épisodes qu'il n'y en a dans les segmentations de référence.

Différents modèles, basés sur l'algorithme de segmentation thématique TextTiling (Hearst, 1997), sont proposées dans cette partie pour tenter de répondre à ces limitations. L'algorithme TextTiling est une méthode de segmentation non supervisée, qui part du postulat que chaque *sous-thème*, également appelé *passage*, se caractérise par un vocabulaire redondant. L'algorithme consiste à analyser l'ensemble des cooccurrences lexicales et à associer les changements de vocabulaire à des changements thématiques. L'implémentation de TextTiling fournie dans NLTK (Loper and Bird, 2002) sert de base aux modèles proposés ci-dessous.

Le comportement par défaut de l'implémentation consiste à utiliser une procédure de normalisation qui associe les frontières thématiques aux frontières de paragraphe les plus proches (**P**). Deux méthodes de normalisation supplémentaires ont été réalisées. La première consiste à ignorer les marques de paragraphe et à considérer chaque frontière de phrase comme un candidat potentiel (**S**). Cette alternative a pour but de permettre la détection des 29% de frontières non associées à

des frontières de paragraphe. Le deuxième critère de normalisation proposé consiste à utiliser les annotations en tour de parole obtenus via des procédures automatiques (**TP**). Cette stratégie a pour but de ne considérer que les candidats détectés par l'automate FpTpAuto et utilise l'algorithme TextTiling pour réduire le taux de fausse détection de frontière, afin d'améliorer la précision du modèle et le ratio de frontières épisodiques (RFE).

Le comportement par défaut de l'algorithme TextTiling consiste à fonder ses décisions sur les cooccurrences lexicales (**W**). Une *stop list* est utilisée pour exclure les mots les plus fréquemment utilisés dans la langue, susceptibles de bruite l'analyse des cooccurrences. Cette particularité a pour effet de ne conserver que les mentions explicites des personnages et de supprimer tous les pronoms. Les changements d'épisodes peuvent être liés à l'introduction d'un nouveau personnage, et à la présence des personnages à un moment donné. En partant de cette hypothèse, une variante de TextTiling a été mise au point, consistant à remplacer toutes les mentions des personnages rencontrées dans le texte (entités personne) par un identifiant unique associé à leur attribut anaphorique. Les identifiants sont ensuite utilisés en combinaison avec les mots non associés à des entités personne (**W+R**). Une dernière variante consiste à ignorer tous les mots et à n'analyser que les cooccurrences d'identifiants anaphoriques (**R**) pour détecter les changements de scène.

Le comportement de l'algorithme TextTiling est influencé par deux paramètres principaux : la taille des pseudo-phrases utilisées (**w**), exprimée en nombre de mots et la taille de bloc (**k**), correspondant au nombre de pseudo-phrases considérées simultanément. Chaque combinaison de paramètres a pour effet d'optimiser ou non une métrique d'évaluation de segmentation. Un ensemble exhaustif de paramètres est évalué, en considérant l'ensemble des combinaisons faisant varier la taille de bloc  $k$  entre 3 et 15 pseudo-phrases, et la taille des pseudo-phrases entre 3 et 30 mots.

Les modèles optimisant respectivement la F-mesure, WindowDiff et Pk apparaissent dans les tables 3.5, 3.6 et 3.7. D'après ces trois métriques, l'utilisation de méthodes de normalisation des frontières combinant les paragraphes et les tours de parole permet systématiquement d'améliorer les modèles de prédictions de frontières. Le meilleur modèle obtenu selon la métrique WindowDiff utilise à la fois les cooccurrences de mots et les références vers les personnages de l'histoire. La F-mesure et Pk sont optimisées par un même modèle utilisant une taille de pseudo-phrase  $w = 5$ , associées à une taille de bloc  $k = 12$ . Le modèle utilise pour ses prédictions les marques de paragraphe, les tours de parole et les références vers les personnages, en ignorant l'ensemble des mots. Ces propriétés appuient la thèse que les algorithmes de segmentation non-supervisée peuvent être améliorés en remplaçant les entités lexicales par des unités linguistiques de plus haut niveau, sous réserve que des procédures d'extraction automatique de ces unités soient disponibles.

Aucun des modèles de segmentation non-supervisée présentés dans cette partie ne permet d'obtenir une meilleure F-mesure sur la tâche de segmentation que l'automate FpTpAuto présenté dans la sous-section précédente. Néanmoins, ces modèles permettent améliorer la qualité globale des segmentations, estimée à l'aide de l'ensemble des métriques.

TABLE 3.5 – Évaluation de modèles de segmentation en épisodes basés sur TextTiling et optimisant la F-Mesure

Norm	Features	params	RFE	Prec.	Rec.	FMes.	WD	Pk
S	W	(5, 4)	2.72	0.15	0.41	0.22	0.76	0.55
S	R	(5, 3)	2.57	0.16	0.40	0.22	0.74	0.54
S	W+R	(2, 3)	3.33	0.16	0.53	0.25	0.82	0.54
P	W	(6, 5)	1.40	0.34	0.48	0.40	0.51	0.41
P	R	(12, 5)	1.40	0.37	0.52	0.44	0.47	0.37
P	W+R	(15, 3)	1.48	0.34	0.51	0.41	0.51	0.41
TP	W	(12, 5)	1.11	0.42	0.47	0.44	0.46	0.38
TP	R	(12, 5)	1.19	0.44	0.52	0.48	0.43	0.34
TP	W+R	(14, 3)	1.34	0.42	0.56	0.48	0.46	0.36

TABLE 3.6 – Évaluation de modèles de segmentation en épisodes basés sur TextTiling et optimisant WindowDiff

Norm	Features	params	RFE	Prec.	Rec.	FMes.	WD	Pk
S	W	(2, 28)	0.53	0.17	0.09	0.12	0.48	0.45
S	R	(3, 29)	0.49	0.18	0.09	0.12	0.46	0.44
S	W+R	(1, 29)	0.49	0.16	0.08	0.11	0.46	0.43
P	W	(2, 28)	0.51	0.39	0.20	0.26	0.43	0.40
P	R	(3, 29)	0.48	0.43	0.20	0.28	0.42	0.39
P	W+R	(5, 29)	0.46	0.39	0.18	0.24	0.43	0.40
TP	W	(4, 23)	0.61	0.47	0.28	0.35	0.42	0.38
TP	R	(3, 29)	0.47	0.48	0.23	0.31	0.41	0.38
TP	W+R	(1, 24)	0.56	0.47	0.26	0.34	0.40	0.37

TABLE 3.7 – Évaluation de modèles de segmentation en épisodes basés sur TextTiling et optimisant Pk

Norm	Features	params	RFE	Prec.	Rec.	FMes.	WD	Pk
S	W	(2, 28)	0.53	0.17	0.09	0.12	0.48	0.45
S	R	(10, 23)	0.60	0.20	0.12	0.15	0.47	0.44
S	W+R	(1, 29)	0.49	0.16	0.08	0.11	0.46	0.43
P	W	(5, 15)	0.87	0.38	0.33	0.35	0.45	0.39
P	R	(15, 7)	1.27	0.37	0.47	0.41	0.46	0.37
P	W+R	(6, 9)	1.07	0.36	0.39	0.38	0.47	0.40
TP	W	(5, 14)	0.84	0.45	0.38	0.41	0.42	0.37
TP	R	(12, 5)	1.19	0.44	0.52	0.48	0.43	0.34
TP	W+R	(2, 6)	1.23	0.43	0.53	0.47	0.43	0.34

### 3.4.3 Conclusion partielle

Plusieurs modèles permettant de segmenter les contes en séquences d'épisodes ont été proposés dans cette partie. La difficulté de la segmentation en épisodes est en partie liée à l'accord inter-annotateur modéré observé sur le corpus (section ). Il a été montré dans un premier temps qu'un automate utilisant exclusivement les informations liées aux tours de parole et aux marques de paragraphe permet d'obtenir des résultats acceptables. Dans un second temps, une procédure de normalisation utilisant les tours de parole et les marques de paragraphe a été proposée, qui permet d'améliorer systématiquement les performances de l'algorithme TextTiling (Hearst, 1997). Il a également été montré que l'utilisation d'annotations anaphoriques, obtenues manuellement, contribue à améliorer les performances de TextTiling, et que si ces informations sont disponibles, le reste des unités lexicales n'est pas nécessaire pour segmenter un texte. Les résultats obtenus par les différents systèmes restent acceptables, au regard de la difficulté intrinsèque à la tâche.

D'autres tests ont été réalisés, en adaptant les implémentations des algorithmes de segmentation non supervisée TextTiling et C99 (Choi, 2000) fournies dans la suite d'outils MorphAdorner (Burns, 2009) et en utilisant un protocole identique à celui employé dans la section précédente. Des propriétés similaires liées à l'utilisation des marques de paragraphes, des tours de parole et des annotations anaphoriques ont été observées. Cependant, les résultats obtenus étaient moins bons que ceux basés sur l'adaptation de l'implémentation de TextTiling fournie dans NLTK (Loper and Bird, 2002) et le détail de ces expériences n'est pas mentionné dans ce manuscrit.

La limitation principale des modèles proposés ici vient de leur incapacité à détecter des frontières d'épisode non associées à des marques de paragraphe. La gêne occasionnée par cette limitation reste très dépendante de la manière dont les segmentations doivent être utilisées par la suite. Dans un contexte de synthèse de parole, une analyse informelle réalisée sur le corpus a montré que les pauses associées aux frontières d'épisode étaient plus longues lorsque les frontières étaient associées à des marques de paragraphe. Ce défaut de détection des frontières non associées à des paragraphes ne devrait donc pas pénaliser trop lourdement une analyse destinée à la synthèse, comme c'est le cas de celle-ci. L'évaluation de la qualité intrinsèque des segmentations était rendue difficile à cause du nombre de métriques complémentaires à optimiser. Pour cette raison, dans les cas pour lesquels la segmentation n'est qu'un traitement intermédiaire d'une chaîne de traitements plus vaste (par exemple en vue de la synthèse ou du résumé automatique), il semble judicieux d'évaluer la qualité des segmentations en sortie de chaîne, lorsque cette option est possible.

Les améliorations liées à l'utilisation des annotations anaphoriques manuelles relatives aux personnages suggèrent une piste d'amélioration, qui nécessite de mettre au point des systèmes de résolution des coréférences fonctionnant sur les récits. Une autre piste d'amélioration suggérée par cette observation consiste à généraliser les référencements des entités de personnage aux entités spatiales et temporelles. D'autres pistes d'amélioration pourraient passer par la mise au point de procédures spécialisées dans la détection des refrains, qui sont considérés comme des épisodes dans ces expériences, mais qui d'un point de vue structurel, peuvent être assimilés à des sous épisodes



ou à des ponts inter-épisodes. Il est également envisageable d'utiliser des procédures spécialisées pour détecter le début de l'élément déclencher, ou de l'épilogue, en utilisant des règles lexicales.

Dans le cadre de ces expériences, l'étiquetage des épisodes (élément déclencheur, scène) n'a pas été abordé. Celui-ci est effectué dans un second temps de manière empirique, en considérant que le premier segment est le titre, le deuxième est la situation initiale, le troisième l'élément déclencheur, le dernier l'épilogue, et tous les autres sont des scènes.

### 3.5 Module d'extraction des actes de dialogue et de modes de communication

La détection des actes de dialogue (Rosset et al., 2008) et des modes de communication permet de décrire les phrases assimilables à un discours direct. Dans le cadre de la réalisation d'un module de contrôle de la synthèse de parole, ces informations peuvent être utilisées pour décrire les phrases contenues dans les tours de parole attribués aux personnages. Le module nécessite une annotation préalable du texte en parties du discours (section 3.2.2), en tours de parole (section 3.3) et en pseudo-phrases, déduites à partir des tours de parole et du processus de tokenisation (section 2.3.2).

Les actes de dialogue et les modes de communication sont détectés à l'aide de règles, pouvant nécessiter la prise en compte du contexte des phrases. Le moteur d'expressions régulières Wmatch (Galibert, 2009; Rosset et al., 2009) est utilisé afin de définir des règles s'appliquant à des structures arborescentes, telles que les documents XML.

#### 3.5.1 Modes de communication

Quatre modes de communication sont définis pour qualifier les phrases attribuées à des personnages : *pleur*, *cri*, *exclamation* et *rire*. Chaque mode est associé à une liste de verbes expressifs, dont la présence est vérifiée lors du processus d'annotation.

Pour chaque mode, deux dictionnaires sont constitués. Le premier contient l'ensemble des verbes conjugués à la première personne du présent. Ce dictionnaire est utilisé pour détecter les situations pendant lesquelles un personnage donne des informations sur son état au cours d'une citation. Les verbes conjugués peuvent alors être détectés au sein du tour de parole du personnage (cf. exemple 16). Le deuxième dictionnaire contient la forme infinitive des verbes, les participes passé et présent, ainsi que les flexions correspondant aux deuxièmes et troisièmes personnes du présent, du passé simple et de l'imparfait. Il est utilisé pour détecter les passages où le narrateur précise le mode de communication utilisé par le personnage. Les modes peuvent être mentionnés pendant le tour de parole du narrateur précédent (cf. exemple 17) ou succédant le tour de parole du personnage concerné (cf. exemple 18). Le dictionnaire peut également être utilisé dans le cadre de dialogues, lorsqu'un personnage A mentionne le mode de communication d'un personnage B. Cette mention

peut être réalisée au cours du tour de parole précédent (exemple 19) ou suivant (exemple 20) le tour de parole du personnage B.

(16) - je pleure parce que j' ai très peur !!

(17) - " notre mère est morte ; le singe l' a tuée . "  
, répondirent -ils en pleurant .

(18) il s' exclama soudain : - viens le corbeau , je vais devenir ton avocat , on va aller les voir ces volatiles débiles .

(19) - " petits crabes , pourquoi pleurez -vous donc ? "  
- " notre mère est morte ; le singe l' a tuée . "

(20) - Donnes moi ton cheval !  
- Ce n'est pas la peine de hurler, répondit Erik

#### 3.5.2 Actes de dialogue

Un ensemble de cinq actes dialogue sont définis : les *questions*, les *ordres*, les *interdictions* (ordre négatif), les *assertions* et les *négations* (assertion négative). Leur détection est réalisée au niveau de la phrase, sans prendre en compte les phrases avoisinantes.

Les *questions* sont détectées à l'aide de différents marqueurs. Le plus simple consiste à repérer les phrases se terminant par un point d'interrogation. Un autre marqueur spécifique consiste à repérer les verbes suivis par des pronoms personnels ("est-elle", "fais-tu"). Certaines expressions spécifiques sont également utilisées ("est-ce que", "je voudrais savoir"...).

Les *ordres* sont associés aux phrases contenant des verbes conjugués à l'impératif. Pour accroître la robustesse des règles de détection, dépendantes de l'annotation en parties du discours, des règles supplémentaires sont utilisées qui englobent les verbes conjugués à la troisième personne de l'indicatif présent, suivis par un pronom personnel complément (moi, lui, toi). Lorsqu'un *ordre* commence par l'adverbe de négation "ne", la phrase est assimilée à une *interdiction*.

Les *assertions* et les *négations* sont les actes de dialogues utilisés par défaut. La distinction entre les deux actes réside dans la présence de l'adverbe de négation "ne" dans les négations.

#### 3.5.3 Conclusion partielle

Des règles utilisant des expressions régulières ont été implémentées afin de détecter des actes de dialogues et de modes communicatifs, ce qui permet de décrire les phrases attribuées aux personnages. Les règles relatives aux modes communicatifs prennent en compte le contexte immédiat de la phrase considérée et permettent d'extraire des informations au delà du niveau de la phrase. L'évaluation de la pertinence des règles proposées nécessiterait de disposer d'un corpus de référence

pour ces informations. En l'absence de ressource adéquate, les règles seront évaluées en fin de chaîne, sur leur contribution à l'amélioration de l'expressivité d'un système de synthèse de parole.

Parmi les pistes d'amélioration envisageables, l'ajout de catégories de modes de communication tels que les chuchotements permettraient d'enrichir la palette expressive pouvant être réalisée par un synthétiseur. Une autre piste d'amélioration de la procédure de détection de modes pourrait consister à créer de nouvelles règles, non restreintes aux verbes. Le narrateur donne fréquemment des indications sur le mode de communication à l'aide d'adjectifs, utilisés conjointement avec le mot voix : (“avec une grosse voix”, “d’une voix rauque”, “de sa traditionnelle voix tonitruante”, “d’une voix suave”...).

## Chapitre 4

# Conception et Analyse d'un corpus oral de contes

### Sommaire

---

<b>4.1</b>	<b>Corpus oraux de contes annotés . . . . .</b>	<b>60</b>
<b>4.2</b>	<b>Enregistrement d'un corpus de contes . . . . .</b>	<b>61</b>
<b>4.3</b>	<b>Alignement lexical et phonétique de la parole . . . . .</b>	<b>62</b>
4.3.1	Transcription lexicale de la parole . . . . .	62
4.3.2	Procédure d'alignement . . . . .	64
<b>4.4</b>	<b>Annotations . . . . .</b>	<b>66</b>
<b>4.5</b>	<b>Stylisation Prosodique . . . . .</b>	<b>67</b>
4.5.1	Estimation de descripteurs prosodiques bruts . . . . .	70
4.5.2	Stylisation basée sur un modèle de perception tonale . . . . .	71
4.5.3	Enregistrement des données stylisées dans des feuilles de calcul . . . . .	71
4.5.4	Représentation graphique du corpus GV-LEx . . . . .	72
4.5.5	Stylisation basée sur le Prosogram . . . . .	73

---

Ce chapitre traite du corpus audio enregistré et analysé dans le cadre du projet GV-LEx. Ce corpus est composé de 12 contes choisis parmi ceux du corpus texte. Après un bref aperçu de corpus similaires décrits dans la littérature et du type d'information qui en sont extraites (cf. section 4.1), une analyse de ce corpus sera fournie. Les enregistrements sont alignés avec les textes et tout un ensemble d'annotations, décrites dans la section 2.4. Des procédures de stylisation prosodique ont été appliquées au corpus afin de faciliter son analyse et la mise en relation des variations prosodiques et des changements structurels des contes.

## 4.1 Corpus oraux de contes annotés

Les annotations contenues dans les corpus oraux peuvent être regroupées en deux catégories principales : d'une part les annotations basées sur l'analyse du contenu linguistique réalisées à partir d'une transcription textuelle de la parole (type de texte, structures syntaxiques, référencement et descriptions des personnages) ; et d'autre part les annotations permettant une description des variations de ce signal de parole (transcription phonétique, stylisation prosodique, etc.). Le premier type d'analyse est principalement dédié à l'analyse linguistique de l'oral (cf. [Blanche-Benveniste, 2011](#)). Le présent travail se situe dans une optique tenant plutôt de la seconde approche. Il s'agit non pas d'étudier l'oral, mais d'extraire des informations d'un texte lu, dans le but de guider un système de synthèse.

Pour cela, il est nécessaire d'extraire du corpus un ensemble de descripteurs pour étudier leur variabilité. Ces descripteurs peuvent être phonétiques (voir [Adda-Decker et al., 2008](#)), ou prosodique. Dans ce cas, il s'agit d'extraire des informations quant aux structures prosodiques observées dans le corpus et cela demande généralement l'application d'un modèle prosodique. Ce modèle peut être phonologique (par exemple, pour le français : [Mertens et al., 2001](#); [Jun and Fougeron, 2002](#)) et/ou procéder grâce à une stylisation des variations observées (voir par exemple [d'Alessandro and Mertens, 1995](#); [Hirst et al., 2000](#); [Mixdorff and Fujisaki, 2000](#); [Taylor, 2000](#)). Il peut aussi s'agir d'approches plus ciblées, intéressées par un aspect particulier des variations prosodiques, comme la mesure de la proéminence ([Avanzi et al., 2010](#)). Les choix qui sont fait pour ces transcriptions dépendent pour beaucoup du but des analyses et des contraintes liées à l'utilisation de systèmes particuliers.

Les transcriptions sont généralement *alignées* sur le signal de parole, c'est-à-dire associées à des intervalles temporels qui correspondent à leur position sur le signal de parole. L'alignement peut être réalisé manuellement ou via des procédures automatisées. Différentes unités peuvent être utilisées pour effectuer cet alignement : les groupes de souffle, les mots, les syllabes ou les phonèmes.

Un certain nombre de travaux visant à mesurer la variabilité expressive ont conduit à la constitution de corpus oraux de contes annotés (cf. table 4.1). L'intérêt pour les collections de contes lus ou racontés peut s'expliquer par la richesse des phénomènes prosodiques qui y sont observés : des occurrences de discours descriptif, direct ou rapporté, des stratégies d'incarnation de personnages, des réalisations d'affects et d'émotions dans des contextes différents, etc.

On peut observer différentes tendances dans les stratégies utilisées pour observer la variation prosodique dans des contes. Certains auteurs cherchent par exemple à comparer les spécificités prosodiques du *storytelling* à celles d'autres styles (informations, recettes, ...) ([Fackrell et al., 2000](#)). Les travaux de ([Levin et al., 1982](#)) cherchent à mesurer les différences entre les contes *lus* et *racontés* (c'est-à-dire la retransmission orale d'un conte mémorisé préalablement). C'est aussi un aspect du travail de [Mixdorff and Barbosa \(2012\)](#), qui ajoutent la comparaison prosodique entre l'allemand et le portugais brésilien. Des annotations émotionnelles peuvent être utilisées

TABLE 4.1 – Corpus annotés de parole contenant des contes

Name	Language	characteristics
Levin et al. (1982)	Anglais	2 contes , 4 locuteurs
Alm and Sproat (2005)	Anglais	2 contes, 10 minutes
Fackrell et al. (2000)	Néerlandais, Français	Anglais, 6 heures
Theune et al. (2006)	Néerlandais	5 contes (durée entre 5 et 12 mins)
Adell et al. (2005)	Espagnol	1 conte, 798 phrases
Mixdorff and Barbosa (2012)	Allemand, Brésilien	Portugais 1 conte (1500 mots), 2 styles, 15 locuteurs
GVLEX	Français	12 contes, 1 heure

pour identifier les émotions détectées à partir de l'analyse du texte seul (Adell et al., 2005), ou les émotions perçues par les auditeurs lors de l'écoute du signal de parole (Alm and Sproat, 2005). Des annotations structurelles sont utilisées pour décrire les modes discursifs (narration, description, dialogue) (Adell et al., 2005), le suspense, ou la tension croissante (Theune et al., 2006). Une dernière tendance consiste à annoter directement la prosodie du signal à l'aide de structure de haut niveau (Klabbers and van Santen, 2004).

Le corpus oral GV-LEx, décrit dans la suite de ce chapitre constitue un important volume de données enregistrées par un même locuteur. Sa durée est comparable à celle des corpus de contes existant dans d'autres langues mentionnées dans la table 4.1. L'originalité du schéma d'annotation de ce corpus vient de ce qu'il propose des informations linguistiques variées : sur la structuration des contes en épisodes, l'identification et l'attribution des citations directes, des méta-informations de description des personnages, de référencement des mentions des personnages et des entités nommées, ainsi que d'autres informations obtenues à l'aide de procédures automatiques. Les enregistrements sont alignés avec les transcriptions phonétiques du signal, permettant une description fine des propriétés prosodiques observées dans le signal et l'utilisation de modèles de stylisation prosodique.

## 4.2 Enregistrement d'un corpus de contes

Afin de constituer un corpus de référence de ce que peut être la lecture d'un conte dans les conditions décrites pour le projet GV-LEx, douze contes ont été sélectionnés parmi ceux du corpus de texte. Ces contes ont été lus par un locuteur professionnel, habitué aux enregistrements en studio. Les textes normalisés des contes ont été fournis au locuteur avant la séance d'enregistrement afin qu'il puisse les lire à l'avance. Le locuteur a été encouragé à modifier certaines portions de texte s'il jugeait que leur prononciation ne serait pas naturelle. Il a été informé du but de cet enregistrement, à savoir l'analyse prosodique de contes lus et il lui a été demandé de produire les différentes histoires comme s'il s'adressait à un public d'enfants.

Les contes ont été enregistrés dans un studio, pendant une séance qui a duré quatre heures. Les enregistrements résultant durent une heure environ. Le locuteur était assisté par un ingénieur du son avec qui il avait l'habitude de travailler. Les contes ont été enregistrés via un procédé d'*overdubbing*. Ce procédé permet au locuteur d'interrompre l'enregistrement à tout moment, de relire les portions enregistrées en effaçant les parties problématiques. Ce procédé fournit au locuteur un contrôle optimal sur sa prestation et permet d'obtenir des enregistrements contenant moins d'erreurs (répétitions, hésitations, etc.) que s'il s'agissait de parole lue ou spontanée. Cette propriété est très désirable lorsque les enregistrements doivent servir de modèle pour améliorer les capacités expressives des synthétiseurs de parole. Les enregistrements obtenus sont de qualité comparable à celle d'un livre audio, malgré les quelques restrictions nécessaires pour permettre leur utilisation dans le cadre d'analyses prosodiques. Ainsi, la voix n'est pas mélangée à de la musique et aucun effet audio (compression, égalisation) n'a été utilisé. Un seul locuteur a été enregistré. Dans les analyses qui suivent, seule la variation intra-locuteur sera donc étudiée.

Les enregistrements ont été numérisés avec un format d'échantillon flottant de 32 bits, associé à un taux d'échantillonnage de 48 KHz. Ils contiennent deux canaux obtenus à l'aide de deux microphones distincts. Le premier canal correspond aux enregistrements obtenus avec un microphone Neumann U87, généralement utilisés en studio pour enregistrer de la voix ou de la musique. Le deuxième canal a été obtenu via un microphone Earthworks M30, plus adapté aux mesures acoustiques. Le microphone Earthworks a été calibré à l'aide d'un calibreur Brüel et Kjær, permettant d'estimer l'intensité absolue du signal enregistré.

### 4.3 Alignement lexical et phonétique de la parole

L'alignement lexical et phonétique de la parole consiste à déterminer les positions des frontières de mots et de phonèmes dans des enregistrements de parole. Au cours d'une première étape manuelle, les paroles prononcées par le locuteur sont retranscrites, et alignées grossièrement avec le signal de parole. Plusieurs variantes de prononciation peuvent co-exister pour un lexique donné (phonèmes optionnels ou interchangeables). L'étape suivante consiste donc à déterminer la transcription phonétique associée au signal de parole, ainsi puis les frontières des phonèmes sur le signal.

#### 4.3.1 Transcription lexicale de la parole

La transcription lexicale des contes et son alignement grossier avec le signal de parole a été réalisée à l'aide du logiciel Transcriber (Barras et al., 2001). La transcription a été grandement facilitée par l'existence du texte normalisé lu par le locuteur : seules les modifications introduites par le conteur ont été l'objet d'une correction. L'alignement de cette transcription consiste à identifier les frontières des différents *groupes de souffle*, définis comme les passages de parole délimités par des pauses. Une fois les groupes de souffle identifiés, ils sont annotés avec la trans-

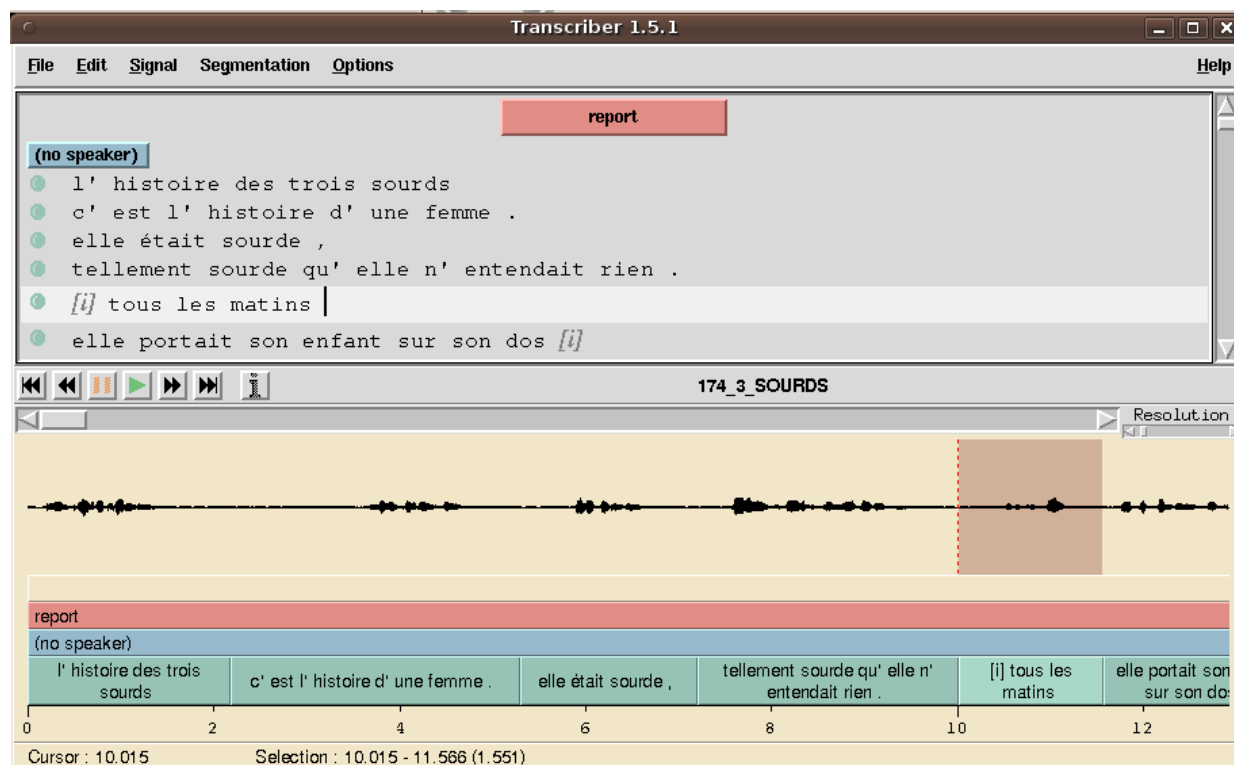


FIGURE 4.1 – Transcription de l’enregistrement d’un conte à l’aide du logiciel Transcriber. Les balises [i] marquent les inspirations.

cription du texte prononcé (figure 4.1). Ils sont ensuite corrigés depuis l’interface de transcription pour prendre en compte les différences entre le script initial et les mots ayant effectivement été prononcés. La transcription des textes est enrichie avec l’annotation des événements non lexicaux observés dans les enregistrements (respirations, rires, claquements de langue, etc.).

Une description sommaire des caractéristiques des transcriptions est présentée dans la table 4.2. Les *pseudo-syntagmes* sont définis comme des groupes de mots séparés par des signes de ponctuation non terminaux. Par exemple la phrase “*Mais son frère, qui était très avare, refusa tout net de l’ aider et le cadet reparti.*” correspond à trois pseudo-syntagmes : “*Mais son frère*”, “*qui était très avare*” puis de “*refusa tout net de l’ aider et le cadet reparti*”. La détection des phrases et des paragraphes, réalisée lors du processus de *tokenisation* (section 2.3.2), est corrigée à l’aide du système d’annotation automatique des tours de parole présenté dans la section 3.3. Les erreurs de *tokenisation* sont liées à la mauvaise détection d’un certain nombre d’incises, qui sont assimilées à des frontières de phrase ou de paragraphe. Leur correction consiste à détecter un tour de parole de personnage suivi par un tour de parole du narrateur (à l’aide des modèles à base de CRF), puis à vérifier si le tour de parole du narrateur correspond à une incise en utilisant les règles syntaxiques utilisées par l’automate TPAutoPOS, défini dans la section 3.3.1. La pertinence des corrections apportées de cette manière a été validée par une inspection manuelle.

Le sous-ensemble de contes sélectionnés pour être enregistrés, et présenté dans la table 2.2, se



TABLE 4.2 – Caractéristiques des transcriptions des douze contes contenus dans le corpus de parole. Le signe \* indique les mesures effectuées après la correction automatique des erreurs de *tokenisation* liées à la détection des incises. La taille des phrases est mesurée en nombre de mots.

	Moyenne	Minimum	Maximum	Total
Nb Mots	805,3	626	1031	9664
Nb signes de ponctuation	179,4	92	281	2153
Nb pseudo-syntagmes*	129,8	80	215	1157
Nb phrases	70,8	43	127	849
Nb phrases*	64,5	43	122	774
Nb paragraphes	20,9	11	34	251
Nb paragraphes*	19,9	10	33	239
Taille moyenne des phrases	12,4	7,3	20,5	
Taille maximum des phrases	40,3	25	62	

TABLE 4.3 – Différences observées entre les textes des contes fournis au locuteur (9664 mots) et les transcriptions des enregistrements du corpus de parole

Suppressions	Insertions	Modifications	Déplacements
40	31	41	4

caractérise par des textes légèrement plus longs que la moyenne du corpus texte (805 mots contre 752). La taille moyenne des phrases varie du simple au triple en fonction du conte considéré ce qui dénote une certaine diversité syntaxique au sein du corpus oral.

Les différences observées entre les textes originaux et les transcriptions sont décrites dans la table 4.3, en termes de nombre de suppressions ou d'insertions de mots, de modifications (ce qui inclus le remplacement par un synonyme) et de changements de position d'un mot au sein d'une phrase, réalisés par le locuteur. Les faibles taux de différences observées (de l'ordre du pourcent) montrent une fidélité au texte original raisonnable. L'analyse manuelle de ces différences montre que les modifications de script réalisées par le locuteur n'affectent ni la structure des contes, ni la structure des phrases.

### 4.3.2 Procédure d'alignement

L'alignement lexical et phonétique est effectué à l'aide d'une procédure semi-automatique qui permet d'obtenir les frontières des mots, des phonèmes, et des pauses. Différents types de pauses, telles que définies dans (Campione and Véronis, 2004), sont identifiés. Les *pauses silencieuses* correspondent aux portions de parole pour lesquelles la production vocale est interrompue. Cette définition inclue les phénomènes de respirations audible, qui sont identifiés à l'aide de symboles spécifiques. Les *pauses remplies*, également appelées *pauses sonores*, correspondent à des portions de parole associées à des éléments quasi-lexicaux, tels que des marques d'hésitations (euh), exclamations, grelottements, ...

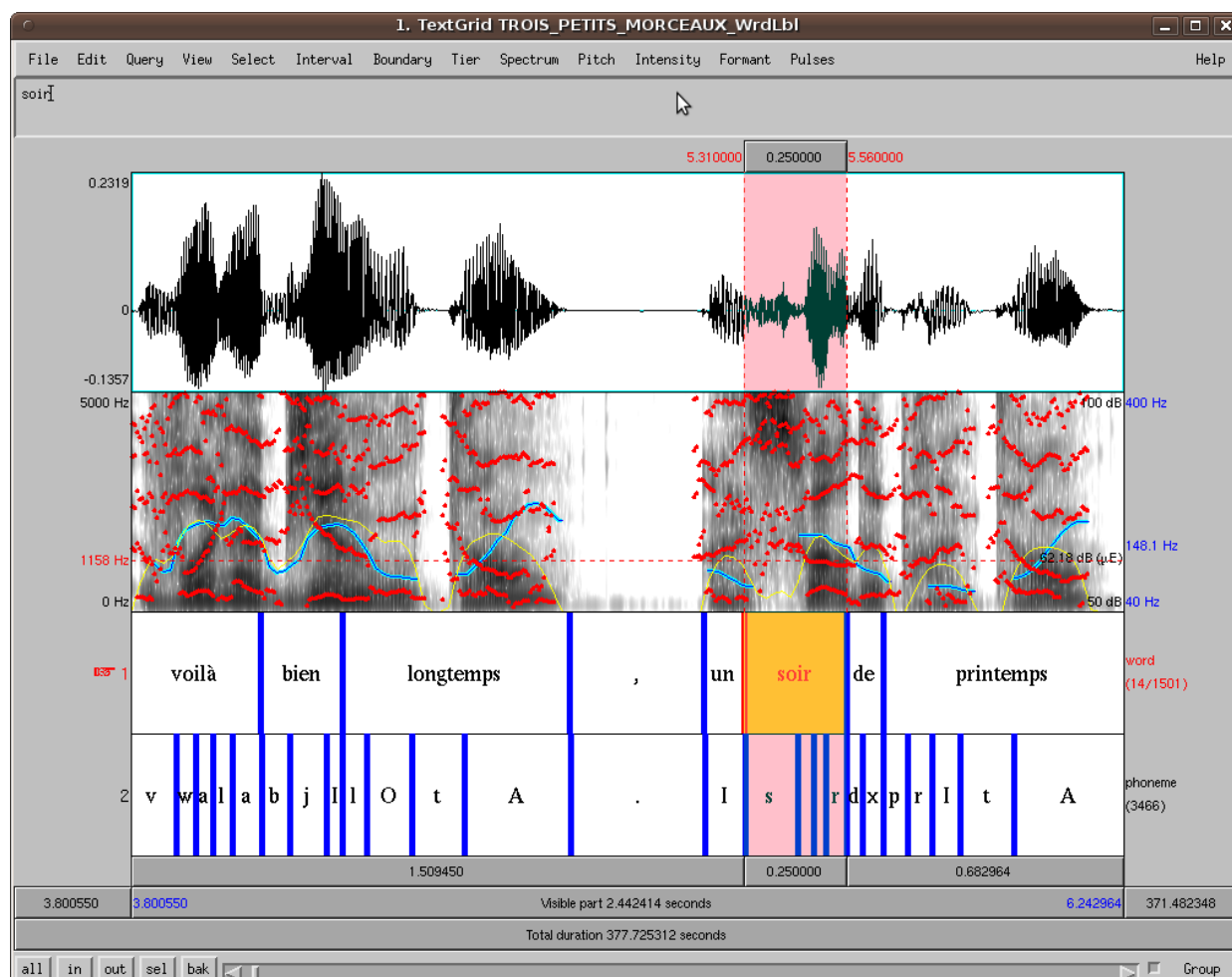


FIGURE 4.2 – Visualisation et édition des frontières lexicales et phonétiques depuis Praat

Outre la réduction du coût de l'annotation, l'utilisation de systèmes semi-automatique permet d'obtenir des positions de frontières phonémiques reproductibles et cohérentes sur l'ensemble du corpus. Sachant que les limites des unités phonémiques sont souvent issues de choix ne pouvant être considérés comme une vérité incontestable (Astesano et al., 1995). Le système d'alignement retenu prend en entrée une segmentation préalable des transcriptions en groupes de souffle, alignées avec le signal de parole (segmentation décrite dans la sous-section précédente) et nécessite l'utilisation de fichiers audio échantillonnés à 16 KHz. Le dictionnaire de variantes de prononciation décrit par Adda-Decker and Lamel (1999) est utilisé conjointement avec le système d'alignement automatique du LIMSI (Gauvain et al., 2005), chargé de déterminer la transcription phonétique la plus plausible pour un segment de parole, ainsi que les frontières des phonèmes et des mots.

Les procédures d'alignement automatique permettent d'obtenir des segmentations relativement fiables, à l'exception des positions exactes des frontières des groupes de souffle, qui ont été corrigées manuellement. L'édition manuelle des frontières des groupes de souffle a été réalisée grâce au logiciel Praat (Boersma, 2002). La figure 4.2 illustre l'interface d'édition de ce logiciel, qui permet de visualiser simultanément les frontières des mots et des phonèmes, le signal audio et sa

TABLE 4.4 – Description des données obtenues suite à l’alignement lexical et phonétique du corpus de parole (durée totale : 3593 secondes)

Type	min	max	moyenne	total	couverture (secondes)
Phonèmes	1995	2914	2373,8	28486	2531,4
Voyelles	894	1316	1077,4	12929	1199,6
Pauses silencieuses	81	187	120,8	1449	837,7
Respirations audibles	23	116	56,9	683	203,8
Pauses remplies	1	33	11,5	138	20,4

décomposition spectrographique, superposée à l’estimation de la fréquence fondamentale du signal (de couleur cyan), à son intensité (en jaune) et ses formants (rouge). L’ensemble des frontières est inspecté manuellement dans un second temps et seules les erreurs flagrantes (de phonétisation, de positionnement des frontières, de pauses remplies) ont été corrigées. Toutes les annotations sont finalement enregistrées dans des fichiers au format TextGrid de Praat. Les statistiques obtenues à l’issue de cette procédure d’alignement sont fournies dans la table 4.4.

## 4.4 Annotations

Toutes les annotations manuelles obtenues lors de la constitution du corpus texte (section 2.4) ont été adaptées manuellement pour correspondre aux transcriptions du signal de parole. Les procédures d’enrichissement automatique des unités lexicales, d’annotations des actes de dialogue et des modes expressifs (cf. sections 3.3.2 et 3.5) ont été utilisées pour annoter les transcriptions du signal de parole. Des procédures automatiques supplémentaires sont utilisées afin d’annoter des unités d’analyses prosodiques courantes à partir des transcriptions phonémiques. La procédure de syllabation décrite par Adda-Decker et al. (2005) a été utilisée pour obtenir les frontières des syllabes. Les *Groupes Inter Perceptual Center Group* (GIPC) ont aussi été extraits à partir de la chaîne phonémique en utilisant l’approximation utilisée par Barbosa and Bailly (1994) qui consiste à faire débiter un GIPC à chaque noyau vocalique (sauf dans le cas des consonnes suivant immédiatement une pause). Les annotations relatives aux tours de parole des personnages sont enrichies manuellement afin d’inclure des méta-informations décrivant les caractéristiques globales des personnages des contes. Ces informations sont détaillées dans la table 4.5.

La table 4.6 décrit la couverture des annotations linguistiques dans le corpus de parole. Le corpus de parole contient une plus grande proportion de citations directes (40% de texte attribué aux personnages) que le corpus textuel (cf. table 4.6) et permet donc de fournir des données optimisées pour l’étude du discours narratif. On peut ainsi s’attendre à observer plus de variations dans les stratégies de discours direct rapporté, en ce qui concerne l’expressivité au niveau de la phrase comme pour les phénomènes liés aux caractéristiques globales des personnages incarnés.

Toutes ces annotations sont alignées avec le signal de parole et enregistrées dans des fichiers au format TextGrid, en utilisant une *tier* par type d’information annotée. La figure 4.3 représente une

TABLE 4.5 – Méta-informations utilisées pour l’enrichissement de la caractérisation des personnages du corpus de parole

information	attributs
âge	enfant, jeune, adulte, vieux
Genre	masculin, féminin
Taille	petit, moyen, grand, très grand
Type	humain, fée, loup, élément naturel, ...
Valence	gentil, méchant
héro	vrai, faux
Catégorie proppienne (Propp, 1928)	agresseur, auxiliaire, donateur, faux héros, héros, mandateur, objet
Rôle actantiel (Greimas, 1966)	sujet, objet, destinataire, destinataire, adjuvant, opposant

annotation simplifiée. Les différentes *tiers* présentes sur la figure correspondent, de haut en bas, aux épisodes, tours de parole, pseudo-phrases, genre des personnages, valence des personnages, taille des personnages, entités personne, segmentation en paragraphes, segmentation en phrases normalisées, segmentation en mots, parties du discours, phonèmes, GIPC, syllabes, et signes de ponctuation. Le format TextGrid a été retenu pour le stockage dans le but de faciliter la diffusion du corpus : il permet de réaliser des analyses prosodiques sous Praat et est pris en charge par de nombreux logiciels. Il est de plus aisément convertible dans d’autres formats de fichier.

## 4.5 Stylisation Prosodique

Le *Trésor de la Langue Française* (Dendien and Pierrel, 2003) propose plusieurs définitions du terme *stylisation*, et donne comme synonymes *schématisation* et *simplification*. Dans le domaine des arts plastiques, il s’agit du “*fait de représenter un objet en le réduisant à ses caractères les plus typiques ou en lui donnant une configuration conventionnelle*” Il s’agit donc par analogie du “*fait de dégager des lignes principales*”. Dans le domaine littéraire, ce terme correspond au “*fait de s’exprimer de manière concise, en se limitant aux traits essentiels, en estompant les détails trop particuliers*” et donc au sens figuré : “*fait de se limiter aux éléments les plus caractéristiques ou conventionnels*”.

La prosodie peut être décrite comme la perception des grandeurs de hauteur, de longueur et de sonie, ayant comme corrélats acoustiques les plus classiques la fréquence fondamentale (F0 en Hz ou demi-tons), la durée (en seconde) et l’intensité (en dB) (Di Cristo, 2004). On peut rajouter à ces paramètres la notion complexe de “qualité de voix” (voir d’Alessandro, 2006). La *stylisation prosodique* consiste à extraire de ces descripteurs prosodiques les composantes perceptibles intervenant dans les processus de communication (Hermes, 2006). La majorité des modèles de stylisation existants consistent à décrire les contours mélodiques du signal (soit en général sa fréquence fondamentale) à l’aide d’unités discrètes : des segments (t Hart, 1976; d’Alessandro and Mertens,

TABLE 4.6 – Couverture des annotations linguistiques pour les douze contes du corpus de parole

Information	Nb. observations	Couverture (%)	Taille médiane	Taille moyenne	Taille maximum
titre	12	0,6	4	46	8
situation initiale	12	10,4	52	83,8	318
élément déclencheur	12	5,6	36	45,5	119
scène	71	73,2	83	99,6	383
refrain	33	5,5	12	16,1	70
épilogue	12	4,7	30	37,6	77
Tours de parole du narrateur et des personnages					
personnages	304	39,8	8	12,6	87
narrateur	236	60,2	10	24,7	369
Éléments semi-structuraux et lexicaux					
énumération	146	16,3	8	10,8	29
entité nommée spatiale	162	4,8	3	2,9	9
entité nommée temporelle	138	3,4	2	2,3	7
entité nommée personne	1531	26,2	1	1,7	18

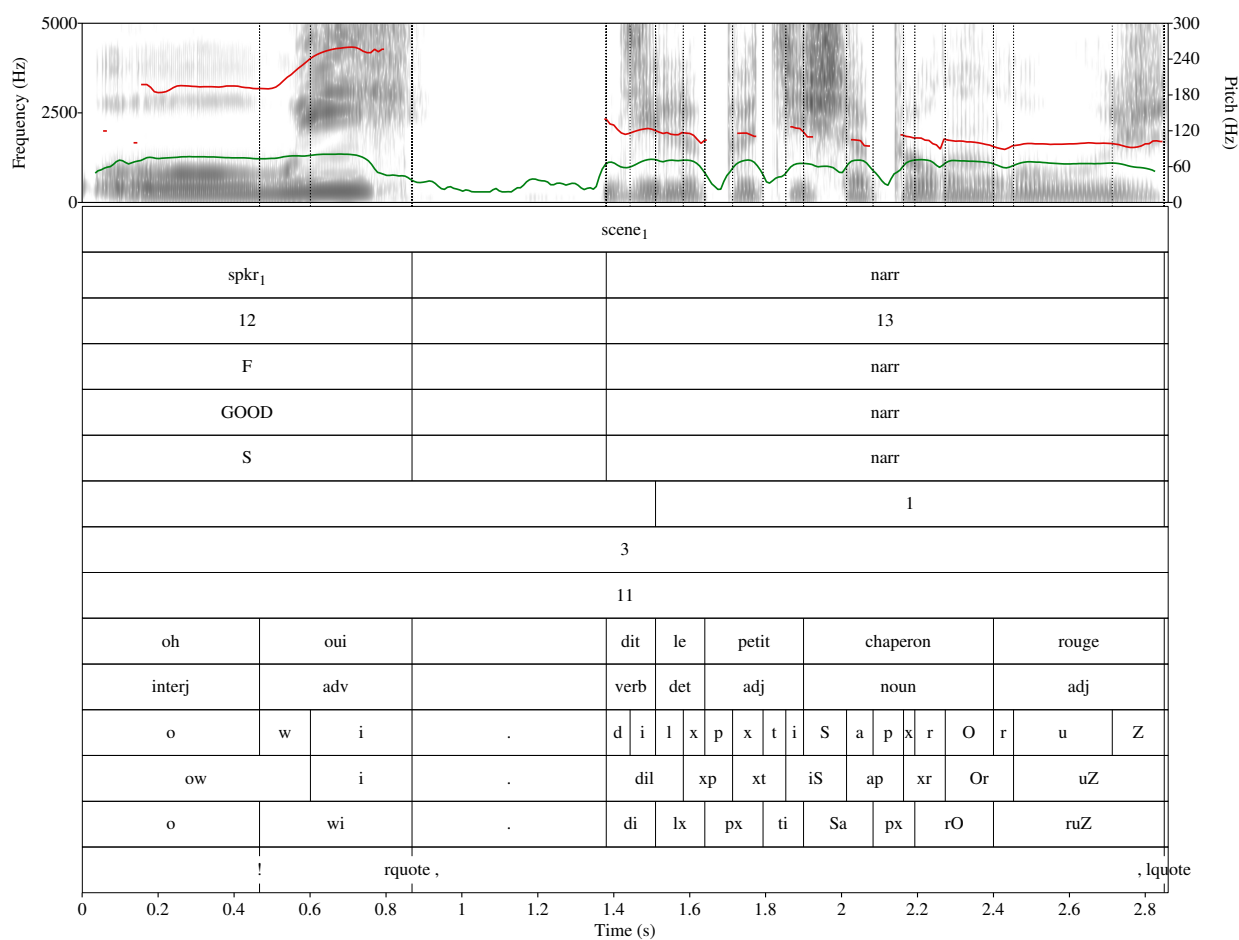


FIGURE 4.3 – Exemple réduit d’annotations alignées avec le signal de parole et visualisées à l’aide du logiciel Praat.

1995), des points cibles (Hirst and Espesser, 1993), ou encore un codage symbolique (Hirst and Di Cristo, 1998). L'intensité et la périodicité du signal sont généralement utilisées implicitement dans ces représentations, afin de déterminer la validité des estimations de fréquence fondamentale.

La procédure de stylisation proposée dans cette section consiste à représenter de manière explicite et concise les mouvements associés à trois descripteurs prosodiques : la fréquence fondamentale, l'intensité et la périodicité.

#### 4.5.1 Estimation de descripteurs prosodiques bruts

L'implémentation Matlab de l'algorithme Yin (De Cheveigné and Kawahara, 2002) est utilisée dans ce travail pour estimer la fréquence fondamentale, l'intensité et la périodicité du signal de parole. L'algorithme est paramétré pour détecter des fréquences fondamentales comprises entre 40 et 1000Hz, estimées sur des fenêtres de 25ms espacées de 2ms chacune. Par analogie avec le système de notation musical, la fréquence fondamentale est convertie en demi-tons relatifs à la fréquence 1Hz (équation 4.1), tel que la note *la* ayant pour fréquence fondamentale 110 Hz corresponde à 81,38 demi-tons.

$$f0_{ST} = 12 * \log_2(f0_{Hz}); \quad (4.1)$$

avec  $f0_{Hz}$  l'estimation de la fréquence fondamentale en Hertz,  $f0_{ST}$  sa conversion exprimée en demi-tons relatifs à 1 Hertz.

L'intensité est exprimée en décibels (dB) et obtenue via l'équation v4.2.

$$I_{dB} = \max(-35, 10 \log_{10}(\frac{I}{I_{Ref}})) \quad (4.2)$$

Avec  $I$  et  $I_{Ref}$  les intensités instantanées estimées sur des fenêtres utilisées pour le calcul de la fréquence fondamentale.  $I$  étant l'intensité estimée sur la fenêtre courante et  $I_{Ref}$  l'intensité maximale estimée sur le corpus. Les estimations d'intensité obtenues via ce procédé sont comprises entre 0 et  $-\infty$ . Un seuil est fixé à  $-35dB$ , en dessous duquel on considère que le signal contenu dans une fenêtre ne correspond pas à de la parole audible, et où les descripteurs prosodiques associés ne sont pas pris en compte par les modèles de stylisation.

L'estimation de la périodicité est généralement utilisée pour mesurer la fiabilité de l'estimation de la fréquence fondamentale sur une fenêtre donnée. Cette estimation peut aussi caractériser certains aspects de la qualité vocale. La mesure de périodicité est obtenue à partir de l'équation 4.3.

$$Periodicite = \min(0, 1 - Ap_{Yin}) \quad (4.3)$$

avec  $Ap_{Yin}$  le coefficient d'apériodicité calculé par Yin et compris dans l'intervalle  $[0, +\infty]$ . Selon

les auteurs de l'algorithme Yin (De Cheveigné and Kawahara, 2002), les estimations de fréquence fondamentale doivent être considérées comme *incertaines* lorsque la mesure de périodicité est inférieure à 0,6, *bonnes* lorsque la mesure est comprise entre 0,6 et 0,8, et *optimales* lorsqu'elles sont comprises entre 0,8 et 1.

#### 4.5.2 Stylisation basée sur un modèle de perception tonale

Une procédure de stylisation basée sur le modèle de perception tonale décrit dans (d'Alessandro and Mertens, 1995) a été développée. La hauteur et sonie perçue sont estimées pour chaque voyelle puis représentées à l'aide d'un ou plusieurs segments. Les bornes de chaque segment sont définies par un triplet correspondant à sa position temporelle, sa fréquence fondamentale et son intensité.

Un seuil de périodicité minimal (égal à 0,2) et un seuil d'intensité minimale (égal à  $-32.5dB$ ) sont définis. Les frontières des voyelles sont réajustées de telle sorte que les fenêtre de signal correspondant au début et à la fin de la voyelle soient associées à une périodicité et à une intensité supérieure ou égale aux seuils. Lorsque cette condition est vérifiée, il est toléré que des fenêtres de signal comprises entre les frontières réajustées ne respectent pas ces conditions. Lorsqu'aucune fenêtre comprise entre les frontières ne respecte ces conditions, le modèle ignore la voyelle.

Les estimations de fréquence fondamentale et d'intensité comprises entre les frontières réajustées sont filtrées par un modèle de moyenne temporelle pondérée<sup>1</sup> (d'Alessandro et al., 1998), défini par l'équation 4.4.

$$WTAM_f(t) = \frac{\int_0^t \exp^{-22(t-\tau)} f(\tau) d\tau}{\int_0^t \exp^{-22(t-\tau)} d\tau} \quad (4.4)$$

Avec  $f$  le descripteur à filtrer, pouvant être la hauteur (en Hertz) ou l'intensité (en décibels).

Le modèle de perception tonale est utilisé avec un seuil différentiel de changement de hauteur égal à 12 demi-tons par seconde, qui permet de fusionner les segments mélodiques de pente inférieure au seuil. Le seuil de glissando n'est pas utilisé, la différenciation entre les tons statiques et les tons dynamiques peut néanmoins être déterminée lors de phases d'analyse ultérieures. Les segments dont la pente est supérieure à 100 demi-tons par seconde sont considérés comme associés à des erreurs d'estimation de la fréquence fondamentale et sont exclus des données. Les bornes des segments, calculées à partir de la fréquence fondamentale filtrée, sont utilisées pour décrire l'intensité filtrée par le modèle de moyenne pondérée.

#### 4.5.3 Enregistrement des données stylisées dans des feuilles de calcul

Les informations prosodiques extraites selon ce procédé sont enregistrées dans des feuilles de calcul, qui seront diffusées avec le corpus, et utilisées dans le cadre d'analyses prosodiques ultérieures.

---

1. Weighted Time Average Model : WTAM



Ces feuilles de calcul contiennent un enregistrement par voyelle, associé à différents champs donnant des informations sur les propriétés acoustiques du signal, telles que :

- Les positions temporelles des frontières des vocaliques, obtenues suite à la procédure d'alignement semi-automatique.
- Les positions temporelles rectifiées automatiquement pour être associées à une périodicité supérieure à 0,2, une intensité supérieure à -32,5 dB et ne contenant pas de segments dont la pente excède 100 demi-tons par secondes.
- Les valeurs moyennes, médianes, maximales et minimales de la fréquence fondamentale, de l'intensité et de la périodicité, mesurées à l'intérieur des frontières rectifiées automatiquement.
- Les données obtenues grâce au modèle de perception tonale. À savoir : les positions des différents segments et les valeurs filtrées à l'aide du modèle de moyenne temporelle pondérée et correspondant à la hauteur tonale et à l'intensité mesurées aux frontières des segments.

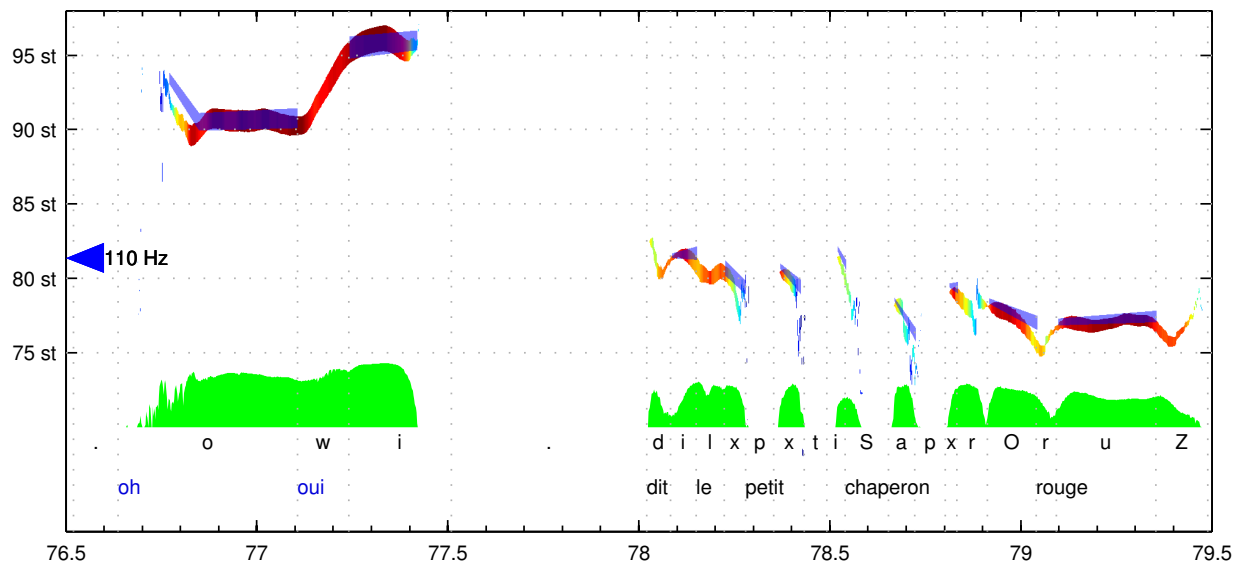
#### 4.5.4 Représentation graphique du corpus GV-LEx

Les descripteurs prosodiques (section 4.5.1) et les segments obtenus à l'aide du modèle de perception tonale (section 4.5.2) sont combinés pour générer des représentations graphiques. Ces représentations permettent de rendre, sur une même figure, la durée, la fréquence fondamentale (brute et stylisée), l'intensité, la périodicité, avec les transcriptions lexicales et phonétiques ainsi que la distinction entre les citations directes et les tours de parole du narrateur. La distinction entre citations directes et tours de parole est effectuée en associant une couleur distincte à chaque personnage. Une telle représentation constitue une sorte de “partition” prosodique et permet de lire les variations expressives observées dans le corpus de contes. Cela permet de guider la modélisation de phénomènes liés à ces variations expressives. Un exemple des représentations graphiques réalisées est représenté à la figure 4.4.

Les descripteurs non stylisés sont représentés à l'aide de courbes affichant la fréquence fondamentale en fonction du temps, courbe dont l'épaisseur est linéairement proportionnelle à l'intensité du signal exprimée en dB, et dont la couleur varie en fonction de la périodicité estimée. On considère que les descripteurs associés à une intensité inférieure ou égale à -35dB ne correspondent pas à de la parole audible. En conséquence, ils sont associés à une épaisseur de trait nulle et ne sont donc pas affichés. La périodicité renseigne sur la fiabilité de l'estimation de la fréquence fondamentale et décrit certains aspects de la qualité de voix. Elle est associée au code couleur “Jet” proposé par Matlab© : le bleu foncé représente les périodicités les plus faibles tandis que le rouge foncé représente les périodicités les plus fortes.

Les segments obtenus à l'aide du modèle de perception tonale sont superposés à cette courbe de F0 et rendus avec une couleur bleue transparente. La largeur des segments est linéairement proportionnelle à l'estimation de l'intensité filtrée.

FIGURE 4.4 – Représentation graphique de la prosodie du corpus GV-LEx, stylisée selon un modèle de perception tonale.

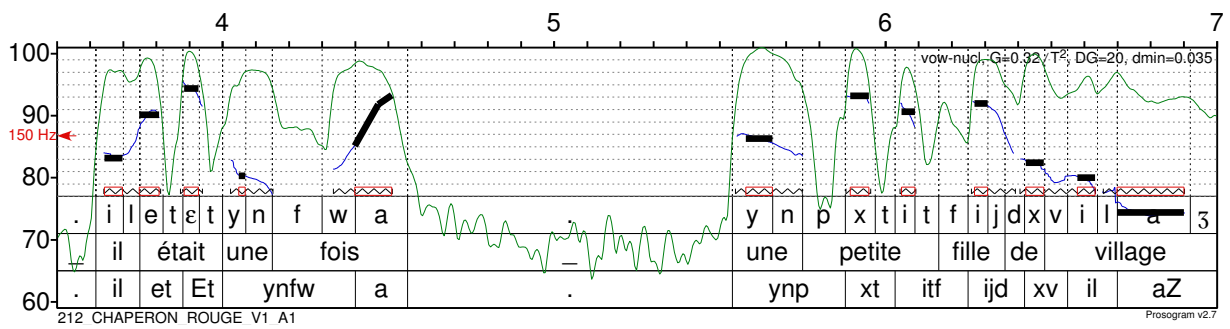


#### 4.5.5 Stylisation basée sur le Prosogram

Le Prosogram (Mertens, 2004) est un logiciel de stylisation prosodique fonctionnant grâce au logiciel Praat Boersma (2002). Le Prosogram fournit une implémentation du modèle de perception tonale décrit dans (d’Alessandro and Mertens, 1995), et montre quelques différences avec l’implémentation proposée dans la section 4.5.2. L’algorithme d’estimation de la fréquence fondamentale et de la périodicité du signal est celui proposé par Boersma (1993). Les seuils qui permettent de déterminer si les fenêtres de signal correspondent à de la parole audible sont calculés différemment et la méthode de rendu de la prosodie stylisée utilise les primitives de dessin disponibles sous Praat (figure 4.5).

Un certain nombre d’analyses prosodiques sur des corpus en langue française ont été réalisées en utilisant des données stylisées à l’aide du Prosogram (Patel et al., 2006; Roekhaut et al., 2010). En conséquence, cet outil est utilisé dans les expériences décrites dans la suite de ce manuscrit, car il fournit une procédure “standardisée” d’extraction de descripteurs prosodiques, ce qui permet de comparer les caractéristiques prosodiques observées sur le corpus GV-LEx aux caractéristiques publiées pour décrire d’autres corpus.

FIGURE 4.5 – Stylisation obtenue avec le Prosogram



## Chapitre 5

# Analyse Prosodique

### Sommaire

---

<b>5.1</b>	<b>Introduction</b>	<b>75</b>
<b>5.2</b>	<b>Descripteurs prosodiques</b>	<b>76</b>
<b>5.3</b>	<b>Comparaison des propriétés prosodiques globales du corpus de contes à d'autres corpus de parole</b>	<b>77</b>
<b>5.4</b>	<b>Propriétés prosodiques liées aux étiquettes d'épisodes</b>	<b>78</b>
<b>5.5</b>	<b>Caractéristiques prosodiques liées à l'incarnation des personnages</b>	<b>80</b>
5.5.1	Impact de l'âge et du genre sur les distributions de fréquence fondamentale et d'intensité	80
5.5.2	Regroupement des personnages basé sur leur profil prosodique	82
<b>5.6</b>	<b>Analyse et modélisation de la durée des pauses</b>	<b>89</b>
5.6.1	Motivations	89
5.6.2	Méthodologie	89
5.6.3	Distributions des pauses	90
5.6.4	Modélisation non paramétrique de la durée des pauses en fonction de la taille du contexte	91
5.6.5	Modèles paramétriques	92
5.6.6	Impact des tours de parole sur la durée des pauses	96
<b>5.7</b>	<b>Conclusion partielle</b>	<b>98</b>

---

### 5.1 Introduction

Ce chapitre présente les analyses prosodiques réalisées sur le corpus GV-LEx. Les analyses décrites dans les sections 5.3 à 5.5 se base sur l'ensemble des paramètres prosodiques décrits dans la section 5.2. La mise en application de ces observations nécessite d'utiliser des synthétiseurs offrant un jeu d'instruction exhaustif. Pour cette raison, un effort particulier est réalisé sur l'analyse de la durée des pauses détaillée dans la section 5.6, permettant formaliser des règles de contrôle pouvant être utilisé avec la majorité des synthétiseurs.

## 5.2 Descripteurs prosodiques

Les propriétés prosodiques du corpus GV-LEx sont décrites tout au long de ce chapitre, à l'aide des descripteurs détaillés ci-dessous :

**NS** : Le Nombre de Syllabes observées sur une portion donnée du corpus

**TP** : Le Taux de Parole, mesuré en nombre de syllabes prononcées par secondes et en excluant les pauses (silencieuses et remplies, cf section 4.3.2) des mesures.

**PTP** : Le Pourcentage de Temps de Pause observé

**NSP** : Nombre moyen de Syllabes entre deux Pauses (taille moyenne des groupes de souffle)

**HES** : Les hésitations, descripteur défini comme le pourcentage de syllabes associées à des hésitations (comme euh, mhhh).

**PVR** : Le Pourcentage de Voyelles Rejetées par le Prosogram (Mertens, 2004, section 4.5.5), généralement associées à un voisement ou à une intensité trop faibles, mettant en cause la fiabilité de l'estimation de la fréquence fondamentale. Ces voyelles sont par la suite ignorées dans les mesures décrivant la hauteur ou l'intensité du signal.

**PTD** : Le Pourcentage de Tons Dynamiques, déterminé à l'aide du Prosogram, correspondant au pourcentage de voyelles pour lesquelles un mouvement mélodique est perceptible

**HM** : La Hauteur Moyenne perçue, exprimée en demi-tons relatifs à 1Hz, obtenue à partir des segments correspondant à la fréquence fondamentale stylisée à l'aide du Prosogram, en ne considérant que la hauteur maximale observée pour chaque voyelle

**IM** : L'Intensité Moyenne, obtenue en ne conservant que l'estimation de puissance maximale observée sur chaque voyelle.

**DHIS** : La Différence de Hauteur Inter-Syllabique, exprimée en demi-tons, obtenue en mesurant la différence de hauteur entre la fin du segment stylisé à l'aide du Prosogram sur la voyelle précédente et le début du segment stylisé pour la voyelle courante.

**DIIS** : La Différence d'Intensité Inter-Syllabique, obtenue en mesurant l'écart entre l'intensité maximale estimée sur la voyelle précédente et la voyelle courante.

**TES** : La TESSiture, exprimée en demi-tons et défini comme l'intervalle compris entre le cinquième et le quatre-vingt quinzième centile de la distribution des hauteurs perçues, obtenue à l'aide du Prosogram

**EI** : L'Étendue des Intensités observées, exprimée en décibels est définie comme l'intervalle compris entre le cinquième et le quatre-vingt quinzième centile de la distribution des intensités maximales observées sur les voyelles

TABLE 5.1 – Comparaison des propriétés prosodiques du corpus GV-LEx et du corpus décrit par Patel et al. (2006)

propriété prosodique	GV-LEx	Patel et al. (2006)
PVR (%)	2	2
TP (syll/sec)	6,15	6,1
PTD (%)	13	2

### 5.3 Comparaison des propriétés prosodiques globales du corpus de contes à d'autres corpus de parole

Cette comparaison vise à caractériser les singularités prosodiques propres à la lecture de contes. Pour cela, les propriétés prosodiques du corpus GV-LEx sont comparées à des analyses similaires réalisées sur d'autres corpus oraux en français - données décrites dans les travaux de Patel et al. (2006) et de Roekhaut et al. (2010).

Les mesures décrites par Patel et al. (2006) sont obtenues sur des actualités, lues par 4 locuteurs différents. Cinq phrases, d'une durée moyenne de 17,3 voyelles sont lues par chaque locuteur et forment un total de vingt phrases d'une durée légèrement inférieure à une minute. Les propriétés de ce corpus sont comparées avec celles du corpus GV-LEx et décrites dans la table 5.1. Le pourcentage de voyelles rejetées et le taux de parole sont similaires pour les deux corpus. La proportion de tons dynamiques observés sur le corpus GV-LEx est plus de six fois supérieure à celle observée sur (Patel et al., 2006). Cette différence importante montre que la lecture de contes est associée à une dynamique mélodique plus importante que celle observée dans la lecture d'actualités. Cette observation est néanmoins nuancée par la faible quantité de données contenues dans le corpus d'actualités.

Les analyses décrites par Roekhaut et al. (2010) portent sur trois styles : des actualités radiophoniques, des discours politiques et des conversations. Pour chaque style, dix minutes de parole sont analysées. Les propriétés de ces styles sont comparées aux propriétés du corpus GV-LEx dans la table 5.2. On y trouvera aussi les spécificités prosodiques propres au narrateur du corpus GV-LEx (sous-corpus *GV-LEx-Narr*). Ce sous-corpus permet une description des propriétés prosodiques associées à la narration séparée des variations prosodiques liées au discours direct rapporté (incarnation de personnages). Le corpus GV-LEx présente un certain nombre de similarités avec le style associé aux discours politiques. Tous deux contiennent un pourcentage de temps de pause important, correspondant à environ 30% de la durée du signal de parole, associé à un nombre de syllabes entre deux pauses environ deux fois moindre que pour les autres styles. Le corpus GV-LEx se distingue des autres styles par une tessiture nettement plus vaste, qui témoigne de la richesse des variations mélodiques propre à la lecture de contes. Le taux de parole du corpus est légèrement plus important que celui observé pour les autres styles, et peut partiellement être expliquées par les conditions d'enregistrement du corpus, permettant au locuteur d'interrompre l'enregistrement à tout moment. Le taux d'hésitations du corpus GV-LEx est relativement faible, mais reste supérieur

TABLE 5.2 – Propriétés prosodiques des corpus GV-LEx et GV-LEx-Narr et des styles correspondant aux actualités radiophoniques (AR), à des discours politiques (DP) et à des conversations (CONV) (Roekhaut et al., 2010)

	AR	DP	CONV	GV-LEx	GV-LEx-Narr
TP (syll/sec)	5,8	4,8	5,3	6,15	6,2
PTP (%)	10,97	31,67	16,73	29,55	24,0
NSP (nb syll)	15	8	16	7,4	7,4
TES (demi-tons)	10,5	10,5	7,4	17,4	16,1
HES (%)	1,83	0,05	7,51	1,01	0,61

à celui observé pour les discours politiques malgré des conditions d’enregistrement permettant plus de contrôle. Ce phénomène peut partiellement être expliqué par la définition utilisée pour annoter les hésitations dans notre corpus, qui englobe à la fois des hésitations involontaires, ainsi que certaines onomatopées non présentes dans les scripts de lecture. 37% des hésitations sont observées dans les tours de parole des personnages, et 27% dans les zones transitoires séparant les tours de parole du narrateur des tours de parole des personnages. Cette observation laisse supposer qu’une partie de ces phénomènes est réalisée volontairement, lors de la prise de parole des personnages. Mis à part ce phénomène, peu de différences sont observées entre le corpus GV-LEx et le sous-ensemble lié au narrateur. L’exclusion des zones de transition entre tours de parole dans le sous-corpus explique les différences observées sur le pourcentage de temps de pause.

## 5.4 Propriétés prosodiques liées aux étiquettes d’épisodes

Cette section présente l’analyse des propriétés prosodiques associées aux différents types d’épisodes dont l’annotation est détaillée dans la section : le titre, la situation initiale, l’élément déclencheur, les scènes, refrains et l’épilogue. Toutes ces propriétés sont décrites dans la table 5.3, en se restreignant au sous-ensemble GV-LEx-Narr défini dans la section 5.3, afin d’exclure de l’analyse les effets liés aux incarnations de personnages, qu’il est plus difficile de relier systématiquement aux épisodes du fait de la répartition de leurs tours de parole. L’analyse séparée des différents épisodes a pour effet d’exclure les pauses inter-épisodes. Ce phénomène, combiné à l’exclusion des pauses séparant les tours de parole intrinsèque au sous-corpus GV-LEx-Narr, se répercute dans les mesures par un pourcentage de temps de pause inférieur à celui observé sur l’ensemble du corpus fourni dans la table 5.2.

L’étiquette épisodique ayant la plus large couverture dans le corpus correspond aux scènes et couvre 63,5% des syllabes observées. Les propriétés prosodiques des scènes correspondent globalement à la moyenne des propriétés observées dans l’ensemble du corpus et se caractérisent par un pourcentage de tons dynamiques un peu plus faible que pour les autres types d’épisodes.

L’énoncé du titre se distingue des autres parties par une proportion de tons dynamiques beaucoup plus élevée, associée à de plus fortes hauteurs moyennes, intensités moyennes, tessiture,

TABLE 5.3 – Comparaison des propriétés prosodiques associées aux étiquettes d'épisodes annotés dans le corpus GV-LEx

	titre	sit. initiale	evt déclencheur	scène	refrain	épilogue
Nb Syllabes	75	1376	484	5115	473	536
Taux Parole (syll/sec)	5,4	6,3	6,4	6,2	6,3	6,1
Temps Pause (%)	5,0	24,9	21,6	20,1	15,0	20,7
Hauteur Moyenne (demi-tons)	85,1	84,6	83,4	84,2	84,1	83,4
Différence Hauteur Inter-Syll (demi-tons)	4,9	4,1	4,5	3,1	2,8	3,2
Tessiture (demi-tons)	21,0	17,1	16,6	15,7	15,4	15,0
Ton Dynamiques (%)	27,0	16,1	17,2	11,3	12,2	12,1
Intensité Moyenne (dB)	67,2	66,3	64,8	66,1	66,7	64,6
Diff Intensité Inter-Syll (dB)	2,7	2,9	3,1	2,8	2,9	2,6
Étendue Intensité (dB)	13,3	14,9	14,6	15,1	15,0	11,8



différence de hauteur inter-syllabique, ainsi qu’au plus faible taux de parole observé sur le corpus. Le pourcentage de temps de pause très faible observé sur cette partie s’explique bien sûr par la très faible durée de ces passages. Ces propriétés tendent à montrer que le style de lecture qui correspond à l’énoncé du titre se caractérise par un volume plus important, associé à des mouvements mélodiques plus amples que ceux observés par la suite et par une vitesse d’élocution plutôt lente. Le locuteur énonce clairement et distinctement le titre pour attirer l’attention de son auditoire.

La lecture des premiers épisodes des contes, à savoir la situation initiale et l’élément déclencheur, partagent un certain nombre de points communs : une forte proportion de tons dynamiques ainsi qu’une différence de hauteur inter-syllabique élevée. Ces propriétés sont associées à une mélodie plus dynamique, qui peut être destinée, comme lors de l’énoncé du titre, à obtenir l’attention de l’auditoire. Ces deux catégories diffèrent principalement au niveau de leur intensité moyenne, qui est plus faible pour l’élément déclencheur, et peut être attribué à une voix plus intime permettant de provoquer des effets de suspense (Theune et al., 2006).

Les refrains se caractérisent par la plus faible différence de hauteur inter-syllabique, montrant des valeurs proches de la mesure observée pour les scènes et les épilogues. Leur faible pourcentage de temps de pause peut lui aussi être expliqué par leur taille moyenne réduite (16.1 mots) et aussi par le fait qu’ils constituent un ensemble homogène, lié par un motif qui se répète d’un refrain à l’autre.

Les épilogues se caractérisent quant à eux par une tessiture, une étendue d’intensité, une intensité moyenne et un taux de parole plus faible que pour les autres épisodes. Ces propriétés sont associées à une intonation moins dynamique (et donc plus “plate”), indiquant la fin de l’histoire.

## 5.5 Caractéristiques prosodiques liées à l’incarnation des personnages

Cette partie décrit la variabilité prosodique associée aux phénomènes d’*incarnation de personnages*. On parle d’incarnation lorsque le texte lu par le conteur correspond à une citation directe attribuée à l’un des personnages. Le but de cette partie est de décrire les relations entre les caractéristiques des personnages (détaillées dans la table 4.5) et les stratégies prosodiques pouvant être utilisées par le conteur lors de leur incarnation.

### 5.5.1 Impact de l’âge et du genre sur les distributions de fréquence fondamentale et d’intensité

La figure 5.1 obtenue avec la fonction *chplot* de R, présente les distributions de F0 et d’intensité pour les incarnations de personnages regroupés par genre (homme/femme) et par classe d’âge (enfant, jeune, adulte, vieux). La distribution du narrateur correspond aux portions de parole qui ne sont pas attribuées à des personnages.

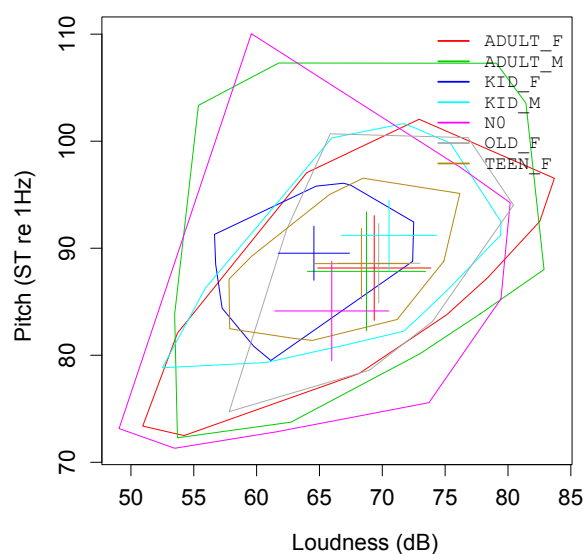


FIGURE 5.1 – Distribution de F0 et d’intensité observées pour les différents rôles narratifs du conteur : narrateur (N0) et différentes catégories de personnages regroupées par âge (enfants, adolescents, adultes, âgé) et par genre (Féminin/Masculin)

On observe que le conteur tend à utiliser une voix plus intense et plus élevée lors des phénomènes d’incarnation de personnages. Cette observation est en accord avec les travaux décrivant les propriétés prosodiques observées dans le discours rapporté direct (Holt, 1996). Elle peut correspondre à une stratégie permettant de rendre plus saillante les voix des personnages.

Pour les classes d’âge jeune, adulte et vieux, on observe des distributions de fréquence fondamentale et d’intensité similaires. Le genre des personnages incarnés a un impact négligeable sur les propriétés de ces distributions. Les registres de voix des personnages de la classe d’âge enfant se distinguent de ceux des autres personnages : notamment par des voix plus aiguës. On observe chez les petits garçons les valeurs médianes de F0 et d’intensité les plus élevées, caractérisées par une différence de 3 demi-tons et 1dB par rapport aux autres classes d’âge. Pour les petites filles on observe une valeur médiane de F0 légèrement plus élevée que pour les autres classes d’âge (1 demi-ton) et intensité sensiblement plus faible (-5dB). Les limites de ces observations sont liées aux différents contextes d’apparition de ces personnages. Cette utilisation d’un registre plus élevé pour des personnages très jeunes (et donc plus faible et plus petits) est cohérente avec la notion de *frequency code* proposé par Ohala (1984), qui montre une corrélation entre la fréquence fondamentale et la taille (et donc la puissance) des mammifères et des oiseaux. Ce lien entre fréquence et taille serait utilisé dans la communication humaine pour signifier des différences de rapport de puissance, notamment. Il pourrait être ici à l’œuvre pour incarner des personnages ayant un rôle plus ou moins empreint de pouvoir.

### 5.5.2 Regroupement des personnages basé sur leur profil prosodique

#### Données

Cette partie consiste à regrouper les personnages incarnés par le conteur en fonction de leur *profil prosodique*. Les profils prosodiques des personnages prennent en compte l'ensemble des descripteurs présentés dans la section 5.2. La valeur moyenne de ces descripteurs est calculée pour chaque personnage, et utilisée pour définir son profil prosodique utilisé dans la suite de cette partie. L'ensemble des profils prosodiques des personnages est détaillé dans l'annexe D. Afin d'obtenir des données fiables, seuls les personnages pour lesquels au moins 50 syllabes sont observées ont été conservés. Le regroupement des profils prosodique est effectué séparément sur les personnages féminins et masculins. Comme la fonction du narrateur ne rentre pas à proprement parler dans cette distinction de genre et que les extraits qui le concernent montrent les variations prosodiques les plus importantes, ils ont été conservés dans les deux ensembles pour servir de référence. Deux personnages correspondant à des oiseaux ont été supprimés car leur genre n'a pas pu être déterminé. Trois autres personnages ont été supprimés car les paramètres prosodiques qui les caractérisent montraient trop de différences avec le reste des personnages (un très grand nombre de dévoisements a été observé pour ces personnages). Les données retenues dans la suite de cette analyse correspondent à 13 profils de personnages féminins, 14 profils de personnages masculins, et au profil du narrateur lorsqu'il n'incarne pas un personnage.

#### Méthode

L'analyse des profils prosodiques est effectuée à l'aide de la suite d'outils *FactoMineR* conçue pour les tâches d'analyse de données exploratoire en R. La procédure *HCPC* (Hierarchical Clustering on Principal Components, [Husson et al., 2010](#)) est utilisée pour obtenir les regroupements et les visualisations des profils prosodiques. Cette méthode est basée sur l'utilisation de procédures d'analyse en composantes principales (ACP, [Shlens, 2014](#)). L'ACP permet de transformer des variables corrélées en nouvelles variables décorrélées les unes des autres. Elle consiste à déterminer un nouvel espace de représentation des données, tel que les dimensions de l'espace d'arrivée correspondent à une combinaison linéaire des dimensions de l'espace de départ. Les axes de l'espace d'arrivée sont perpendiculaires, et déterminés de manière maximiser la variance de la première dimension, puis de la seconde, etc... L'information associée à chaque axe est décrite par son pourcentage de variance. Elle est fréquemment utilisée comme une méthode de réduction de dimensions, en conservant les N premiers axes de l'espace d'arrivée, de manière à ce que la somme des pourcentages des variances des axes retenues soit supérieure à un seuil informatif jugé acceptable. Dans le cadre de notre étude, l'ACP est utilisée comme une méthode de débruitage, en considérant que l'information contenue dans les derniers axes (expliquant le moins la variance) est assimilable à du bruit.

La procédure HCPC est composée de plusieurs étapes successives, détaillées ci-après.

- Appliquer une procédure analyse en composantes principales (ACP) sur les données
- Ne conserver que les N premières composantes de l’analyse pour débruiter les données.
- Utiliser un algorithme de regroupement hiérarchique sur la sortie de l’ACP
- Visualiser les étapes du regroupement sous forme de dendrogramme
- Visualiser les ensembles obtenus dans un espace à deux dimensions, correspondant aux deux premiers axes obtenus à l’aide de l’ACP

Dans le cadre de notre analyse, les 5 (sur 9) premières composantes de l’ACP sont conservées. Le regroupement hiérarchique mené sur la sortie de l’ACP est réalisé en utilisant comme critère d’arrêt le tiers de la distance maximum.

### Personnages féminins

Le groupe féminin comporte 14 personnages (narrateur compris), parmi lesquels on trouve une “maman souris”, des fées, une poule, une mère, une grand-mère et des jeunes filles. L’ACP menée sur le groupe féminin est résumée dans la table 5.4. Les 5 premières dimensions conservées pour la suite de l’analyse permettent d’expliquer 93.75% de la variance des données. Une inspection des deux premières composantes principales de l’ACP menée sur le groupe féminin montre que la première composante principale est liée positivement avec les mesures liées aux variations locales de F0 et d’intensité (comme les glides et les différences inter-syllabiques) et l’intensité moyenne et négativement avec le pourcentage de voyelles dévoisées. La seconde composante principale est liée à la fréquence fondamentale moyenne et à la taille des segments de parole entre deux pauses. Le regroupement hiérarchique permet de distinguer quatre groupes de personnages féminins, représentés sur le dendrogramme de la figure 5.2. L’observation des quatre groupes sur les deux premières composantes principales (cf. figure 5.3), permet d’identifier les stratégies narratives suivantes :

1. Deux jeunes filles ayant un rôle dirigeant sont incarnées par le conteur avec une voix dans un registre élevée, une mélodie plate et un fort pourcentage de voyelles dévoisées.
2. Des femmes adultes ayant un rôle de support proche de celui du narrateur sont incarnées avec une voix assez proche de celle du narrateur.
3. Des femmes plus âgées ou des personnages ayant la charge d’une importante responsabilité (par exemple “maman souris” à la responsabilité d’élever des enfants) sont incarnées avec une intensité moyenne plus élevée et plus de mouvements mélodiques locaux.
4. Un personnage féminin chargé d’aider le héros et incarné avec une voix élevée et intense.

TABLE 5.4 – Données de l'Analyse en Composantes Principale menée sur le groupe de personnages féminins. ctr correspond à la contribution d'une observation pour axe (pour chaque axe  $\sum ctr = 100$ ), cos2 à la qualité de représentation d'une observation sur l'axe considéré

Composantes Principales	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8	Dim.9	
Variance	3.73	2.12	1.21	0.96	0.41	0.27	0.12	0.09	0.07	
% de var.	41.46	23.60	13.46	10.68	4.55	3.05	1.34	1.04	0.82	
% de var. cumulé	41.46	65.06	78.52	89.20	93.75	96.80	98.14	99.18	100	
Individus	Dist	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	cos2
Femme	3.85	0.84	1.34	0.05	3.27	35.87	0.72	1.24	9.11	0.10
Naëlle	3.15	-3.01	17.41	0.92	0.10	0.04	0.00	0.08	0.04	0.00
PCR	4.79	-4.35	36.24	0.82	1.63	8.92	0.12	-0.98	5.64	0.04
Poule	1.81	0.34	0.22	0.03	0.24	0.20	0.02	-1	5.90	0.30
Femme Sourde	2.32	1.58	4.78	0.46	0.91	2.78	0.15	-0.67	2.69	0.08
Mère1	2.32	1.97	7.41	0.72	0.38	0.50	0.03	-0.48	1.34	0.04
Fée1	3.07	-0.87	1.45	0.08	-2.42	19.72	0.62	-0.64	2.43	0.04
Maman Souris	2.40	0.50	0.48	0.04	0.33	0.36	0.02	-1.89	21.06	0.62
Mère2	3.13	-0.57	0.62	0.03	-0.70	1.65	0.05	2.72	43.63	0.75
N0	2.59	-0.47	0.43	0.03	-1.35	6.11	0.27	0.55	1.81	0.05
Fermiere	2.94	2.78	14.77	0.89	0.44	0.65	0.02	-0.16	0.15	0.00
Soeur	2.53	0.14	0.04	0.00	0.73	1.79	0.08	1.01	5.97	0.16
Grand- Mère2	3.27	2.45	11.50	0.56	-1.84	11.42	0.32	0.02	0.00	0.00
Fée2	2.63	-1.32	3.31	0.25	-1.72	9.98	0.43	0.19	0.22	0.00
Variables	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	cos2	
PTD	0.76	15.38	0.57	0.19	1.76	0.04	-0.31	8.06	0.10	
DIIS	0.77	15.93	0.59	0.15	1.06	0.02	0.38	11.89	0.14	
IM	0.74	14.80	0.55	0.55	14.14	0.30	-0.13	1.48	0.02	
NS	-0.20	1.08	0.04	0.79	29.77	0.63	0.27	5.82	0.07	
DHIS	0.83	18.42	0.69	-0.30	4.34	0.09	0.11	1.00	0.01	
HM	0.10	0.29	0.01	0.93	41.19	0.87	-0.16	2.18	0.03	
TES	0.79	16.82	0.63	-0.40	7.61	0.16	-0.19	3.15	0.04	
TP	0.41	4.44	0.17	-0.04	0.08	0.00	0.86	61.25	0.74	
PVR	-0.69	12.84	0.48	0.03	0.04	0.00	0.25	5.17	0.06	

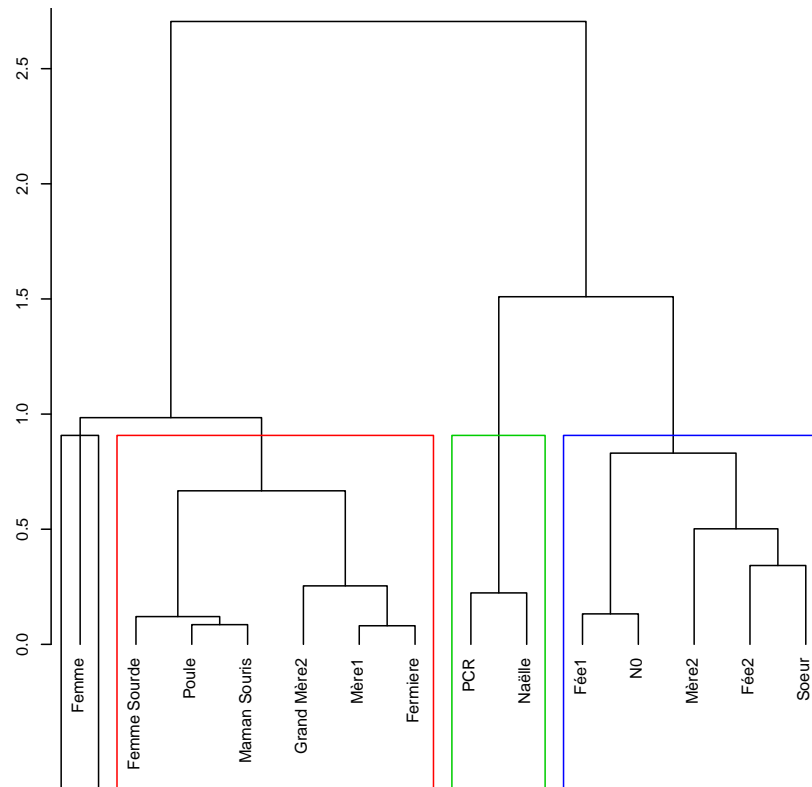


FIGURE 5.2 – Dendrogramme montrant le regroupement hiérarchique des personnages féminins, obtenu à partir de l'ACP menée sur les données issues de la stylisation en Prosogram

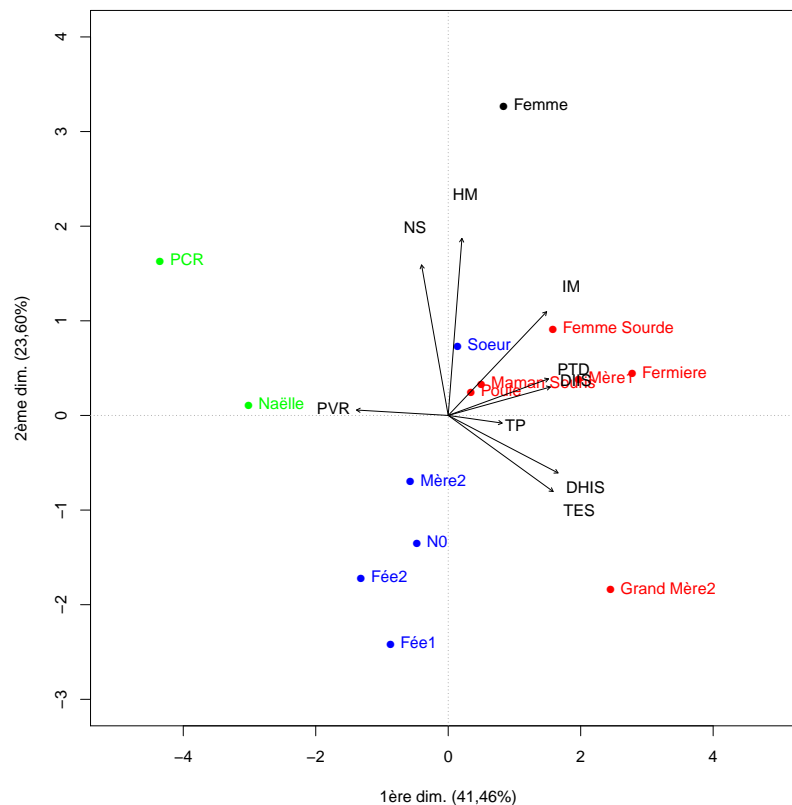


FIGURE 5.3 – Dispersion des personnages féminins sur les deux premières composantes de l'ACP montrant leurs différences en fonction des paramètres prosodiques extraits de la stylisation par Prosogram

## Personnages masculins

Le groupe masculin regroupe 15 personnages (narrateur compris), parmi lesquels on trouve un ours et de nombreux autres animaux, un fermier, un homme sourd, un héros adulte, un garçon ou un petit monstre. L'ACP menée sur le groupe masculin est résumée dans la table 5.5. Les 5 premières dimensions conservées pour la suite de l'analyse permettent d'expliquer 89.6% de la variance des données. L'inspection des répartitions des variables sur les composantes principales montre que la première composante principale est liée négativement à la fréquence fondamentale moyenne, et positivement aux mouvements locaux de F0 (glides et différences inter-syllabiques). La seconde composante principale oppose l'empan de F0 au pourcentage de voyelles dévoisées et à la différence d'intensité inter-syllabique.

Le regroupement mené sur la sortie de l'ACP permet d'identifier cinq ensembles de personnages, représentés par le dendrogramme de la figure 5.4. La dispersion des représentants du groupe masculin sur deux premiers axes de l'ACP est un peu plus complexe (cf. figure 5.5). Les groupes identifiés présentent les caractéristiques suivantes :

1. Des personnages de héros, jeunes et inexpérimentés, incarnés avec une fréquence fondamentale et une intensité moyenne élevée, et peu de variations locales.
2. Des personnages ayant un rôle de support du héros, montrant des caractéristiques prosodiques proches de celles du groupe précédent.
3. Des personnages conseillant et aidant le héros, joués dans un registre bas ou moyen et avec une importante dynamique de F0.
4. Des personnages plus âgés ou plus puissants (prosodiquement proches du narrateur), ils sont incarnés avec un registre bas ou moyen mais avec plus d'intensité et de notables variations prosodiques locales.
5. Un personnage particulier, un ours agressif, ayant une voix très basse, avec d'important changements locaux d'intensité et plus de dévoisements que les autres personnages.

TABLE 5.5 – Données de l'Analyse en Composantes Principale menée sur le groupe de personnages masculins. ctr correspond à la contribution d'une observation pour axe (pour chaque axe  $\sum ctr = 100$ ), cos2 à la qualité de représentation d'une observation sur l'axe considéré

Composantes Principales	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8	Dim.9	
Variance	3.10	1.90	1.51	0.95	0.61	0.40	0.24	0.17	0.12	
% of var.	34.49	21.09	16.76	10.52	6.74	4.48	2.61	1.94	1.37	
% of var. cumulé	34.49	55.57	72.34	82.86	89.60	94.08	96.69	98.63	100	
Individus	Dist	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	cos2
Homme	1.64	0.09	0.02	0.00	0.69	1.69	0.18	1.11	5.47	0.46
Sourd										
Renard	3.51	2.71	15.73	0.60	1.65	9.56	0.22	-0.50	1.11	0.02
Monstre	3.95	-1.75	6.56	0.20	-0.87	2.66	0.05	-3.31	48.31	0.70
Ours	4.88	3.60	27.88	0.55	-3.08	33.26	0.40	-0.02	0.00	0.00
Chien	2.27	-1.07	2.47	0.22	0.61	1.33	0.07	0.39	0.67	0.03
Chef Loup	2.80	0.39	0.33	0.02	-1.78	11.07	0.40	1.78	14.08	0.41
Conseiller	3.53	1.87	7.49	0.28	2.14	16.14	0.37	-0.91	3.62	0.07
Fermier	3.03	-2.00	8.63	0.44	-0.72	1.82	0.06	1.74	13.35	0.33
Maxence	2.40	-1.22	3.20	0.26	-1.15	4.62	0.23	-1.17	6.08	0.24
N0	2.88	1.56	5.22	0.29	-0.55	1.07	0.04	-0.57	1.43	0.04
Veau	3.39	-2.33	11.62	0.47	-0.42	0.63	0.02	0.15	0.10	0.00
Chat	1.98	-1.42	4.35	0.52	0.65	1.48	0.11	0.71	2.26	0.13
Chef	2.62	1.06	2.43	0.16	0.75	1.98	0.08	0.64	1.82	0.06
Erik	1.77	-1.37	4.04	0.60	0.17	0.10	0.01	-0.47	0.97	0.07
Paon	2.56	-0.11	0.03	0.00	1.89	12.59	0.55	0.41	0.73	0.03
Variables	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	cos2	
PTD	0.56	10.05	0.31	0.23	2.72	0.05	0.60	24.25	0.37	
DIIS	0.61	12.13	0.38	-0.52	14.35	0.27	0.32	6.83	0.10	
IM	-0.59	11.40	0.35	-0.06	0.21	0.00	0.52	17.89	0.27	
NS	-0.25	1.95	0.06	0.46	10.98	0.21	0.49	16.11	0.24	
DHIS	0.82	21.73	0.67	0.32	5.51	0.10	-0.01	0.01	0.00	
HM	-0.83	22.14	0.69	0.14	0.97	0.02	-0.10	0.67	0.01	
TES	0.52	8.72	0.27	0.77	31.40	0.60	0.10	0.61	0.01	
TP	-0.45	6.41	0.20	-0.09	0.46	0.01	0.69	31.17	0.47	
PVR	0.41	5.49	0.17	-0.80	33.38	0.63	0.19	2.46	0.04	



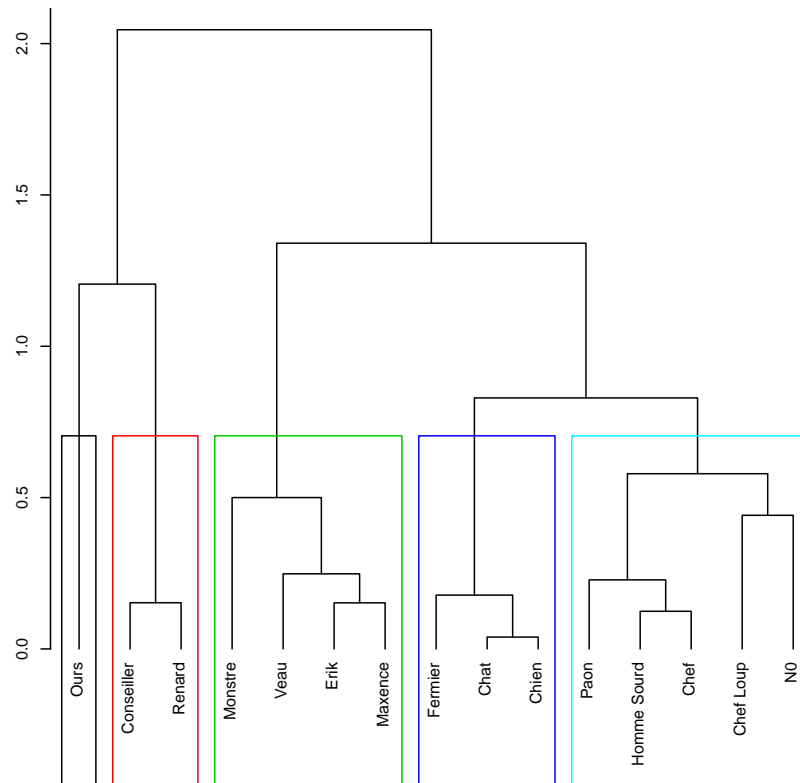


FIGURE 5.4 – Dendrogramme montrant le regroupement hiérarchique des personnages masculins obtenu à partir de l'ACP menée sur les données issues de la stylisation en Prosogram

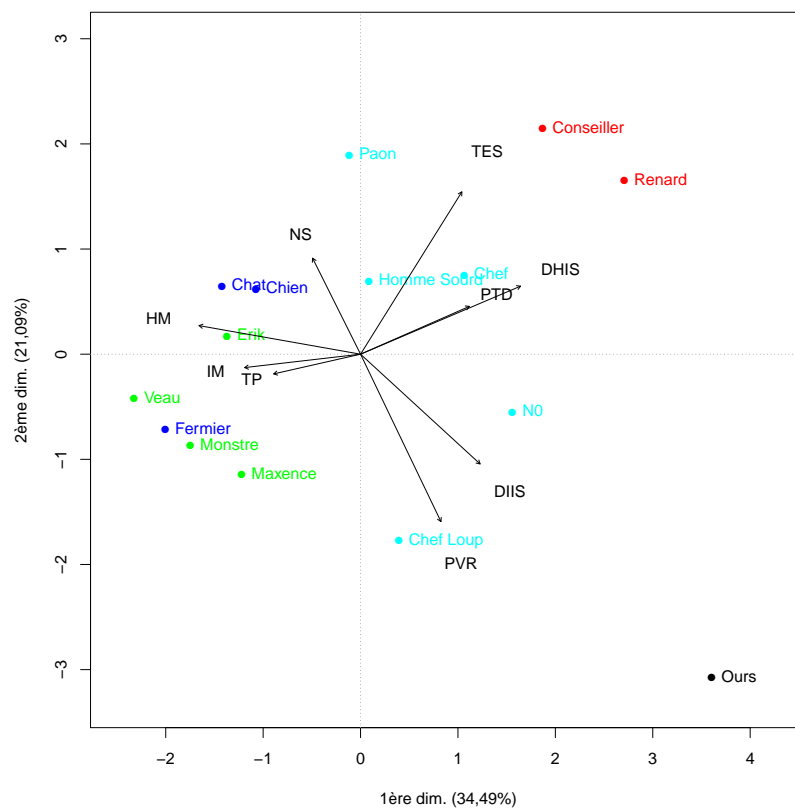


FIGURE 5.5 – Dispersion des personnages masculins sur les deux premières composantes de l'ACP montrant leurs différences en fonction des paramètres prosodiques extraits de la stylisation par Prosogram

## 5.6 Analyse et modélisation de la durée des pauses

### 5.6.1 Motivations

Des études ont montré que le contrôle de la durée et de la localisation des pauses permet d'améliorer l'expressivité de la parole de synthèse (Wang et al., 2008). Le difficile équilibre entre qualité et expressivité de la parole de synthèse (Burkhardt and Stegmann, 2009) est en partie responsable de l'importance des pauses dans ce contexte. Par ailleurs, et comme nous l'avons vu à la section 5.3, les pauses ont une importance particulière dans la lecture de conte, si on la compare à d'autres styles de parole (et ressemble en cela au discours politique, un autre style expressif). La modélisation des pauses dans le cadre de la création d'un modèle de synthèse expressif revêt donc une double importance : il s'agit d'informations vraisemblablement pertinentes pour notre sujet, particulièrement facile à générer sans artefact de synthèse. Étant donné la complexité du contrôle d'un synthétiseur de parole, si l'on ne veut pas faire chuter sa qualité, un effort particulier sera donc fourni pour cette modélisation.

Des études menées sur de la parole neutre produite par des locuteurs non professionnels ont montré des corrélations entre la durée des pauses et la taille des segments de parole qui les environnent (Ferreira, 1991; Zvonik, 2004; Krivokapić, 2007). Nous allons tenter ici de trouver des liens similaires entre la structure des contes et l'apparition et la durée de pauses. À partir de ces observations, des modèles statistiques sur la distribution de ces pauses seront appris, afin de créer un modèle de génération automatique pour le système de synthèse. Les modèles proposés montrent une nette différence entre les pauses apparaissant à l'intérieur des phrases et celles observées entre deux phrases. Le rôle du contexte et de la structure du texte revêt une importance claire : il est donc important de gérer un système de synthèse expressif à partir d'informations au-delà de la phrase. Par ailleurs, comme la longueur des phrases influe sur la durée des pauses, ce paramètre constitue un biais pour mesurer l'influence d'autres facteurs. Des propositions seront donc faites pour normaliser l'influence de la taille des phrases sur la durée des pauses, afin de mesurer plus précisément l'effet des autres facteurs sur cette même durée.

### 5.6.2 Méthodologie

Une pause est définie comme un intervalle entre deux phonèmes d'une durée minimale fixée à 40 millisecondes, correspondant au seuil à peu près maximum de durée de tenue d'une plosive. Toutes les pauses rencontrées dans le corpus GV-LEx ont été inspectées et corrigées manuellement. Dans le cadre de cette étude, elles sont regroupées en quatre catégories principales, mutuellement exclusives : les pauses observées aux frontières des paragraphes (**FPA**), aux frontières des phrases (**FPH**), aux frontières des pseudo-syntagmes (**FPS**, les pseudo-syntagmes sont définis dans la section 4.3.1) et les pauses observées à l'intérieur des pseudo-syntagmes (**IPS**).

Les trois premières catégories de pauses (FPA, FPH, FPS) sont définies à partir des scripts de lecture fournis au locuteur et décrivent des phénomènes associés aux marques de ponctua-

TABLE 5.6 – Distributions des pauses observées dans le corpus GV-LEx

Catégorie de pause	Nombre d'observations	Pourcentage de réalisation	Durée médiane (ms)
intra-pseudo-syntagme	376		261,0
inter-pseudo-syntagme	750	70,9	385,7
inter-phrase	556	98,2	719,9
inter-paragraphe	239	100,0	1126,8

tions (marque de paragraphe, ponctuation terminale, ponctuation non-terminale). On parlera de *réalisation* de pause lorsqu'une marque de frontière est associée à une pause, telle qu'elles sont définies ci-dessus. Les pauses observées à l'intérieur des pseudo-syntagmes ne sont pas associées à des marqueurs de ponctuation.

Ci-dessous, les pauses associées aux frontières de phrases et aux frontières de pseudo-syntagmes seront décrites en fonction de la taille de leur contexte (respectivement les phrases et pseudo-syntagmes précédant et suivant la pause considérée). En d'autres termes, une frontière de phrase (respectivement de pseudo-syntagme) est décrite à l'aide de la taille de la phrase précédant la frontière, combinée à la taille de la phrase observée après la frontière. Ces tailles peuvent être exprimées en nombre de mots, de syllabes, de caractères ou de phonèmes.

Dans un souci d'adéquation avec la perception humaine, les durées des pauses sont converties dans une échelle logarithmique, en utilisant le  $\log_{10}$  de la durée des pauses exprimées en millisecondes (ainsi  $\log_{10}(10 \text{ ms}) = 1$ ,  $\log_{10}(100 \text{ ms}) = 2$  et  $\log_{10}(1000 \text{ ms}) = 3$ ).

### 5.6.3 Distributions des pauses

Les distributions associées à chaque catégorie de pauses sont décrites dans la table 5.6. Le nombre d'observations correspond au nombre de pauses associées à des marques de ponctuation correspondant aux différentes catégories de pauses considérées (FPA, FPH et FPS). Dans le cas des IPS, il s'agit du nombre de pauses observées dans le corpus et non associées à des marques de ponctuation. Le pourcentage de *réalisation* des pauses n'est défini pour les catégories de pauses associées à des marques de ponctuation et correspond au pourcentage de frontières associées à une pause perceptible dans le signal de parole par rapport au nombre de frontières total de ce type. La durée médiane des pauses est calculée en ne prenant en compte que les pauses *réalisées*.

On observe que 29% des frontières de pseudo-syntagmes ne sont pas associées à des pauses perceptibles. Si l'on compare ce nombre au nombre de pauses observées à l'intérieur des pseudo-syntagmes, on observe que 41% des pauses observées à l'intérieur des phrases du corpus ne sont pas associées à des signes de ponctuation, ce qui laisse supposer que le locuteur a effectué une réinterprétation expressive de la structure syntaxique du texte.

La distribution des durées des pauses est aussi illustrée par la figure 5.6. Les distributions associées à chaque type de pause montrent une grande variance ; et si l'on observe un clair accrois-

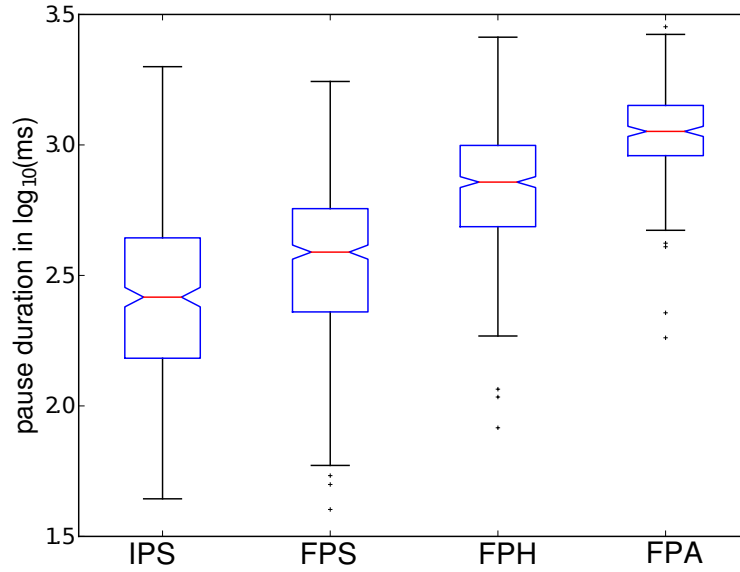


FIGURE 5.6 – Durées des différents types de pauses : à l’intérieur des pseudo-syntagmes, aux frontières des pseudo-syntagmes, des phrases et des paragraphes.

sement de la durée avec les 4 catégories de pauses considérées (selon le schéma :  $FPA > FPH > FPS > IPS$ ), le recouvrement des durées entre deux catégories est très important

L’impact de la longueur du contexte sur la présence de pause aux frontières des pseudo-syntagmes est important : on observe des pauses pour 54% des pseudo-syntagmes dont la taille est inférieure à 12 phonèmes, alors que 87.4% de ces frontières sont associées à des pauses lorsque la taille des pseudo-syntagmes est supérieure à 24 phonèmes. Nous tentons dans la section suivante de modéliser l’influence de la taille du contexte sur la durée des pauses.

#### 5.6.4 Modélisation non paramétrique de la durée des pauses en fonction de la taille du contexte

Une méthode de régression à base de *Kernel* (Nadaraya, 1964), définie par l’équation 5.1, est utilisée pour modéliser la durée des pauses en fonction de la taille de leur contexte.

$$dur(cont) = \frac{\sum_{i=1}^n K(cont, cont_i) dur_i}{\sum_{i=1}^n K(cont, cont_i)} \quad (5.1)$$

La fonction  $dur$  prédit la durée des pauses (en utilisant une échelle logarithmique) à partir de l’argument  $cont = (lp, ls)$  qui est un vecteur bidimensionnel contenant la longueur de l’unité précédente ( $lp$ ) et de l’unité suivante ( $ls$ ). Les prédictions de durée sont basées sur les  $n$  observations de pauses d’une catégorie donnée (inter-phrase ou inter-pseudo-syntagme) présentes dans le corpus, avec  $cont_i = (lp_i, ls_i)$  représentant la taille contextuelle et  $dur_i$  la durée associées à la pause  $i$  du

corpus. La fonction *Kernel*  $K$  permet de pondérer l'impact des données observées dans le corpus en fonction de leur distance avec le point à approximer. Dans le cadre de ce modèle, la fonction *Kernel* choisie est définie par l'équation 5.2, dans laquelle la variable  $h$  représente la largeur de bande du *Kernel* et est utilisée comme un paramètre de lissage.

$$K(x, y) = \exp(-h\|x - y\|^2) \quad (5.2)$$

La durée des pauses est exprimée dans une échelle logarithmique, ce qui ne permet pas de prendre en compte les pauses non réalisées dans les modèles (une durée nulle correspondant à  $-\infty$  à l'échelle logarithmique). Le compromis choisi pour pouvoir prendre en compte la non-réalisation des pauses consiste à pré-traiter les données de telle sorte que les pauses non réalisées soit équivalentes à des pauses très courtes, d'une durée de 40 millisecondes (seuil de définition des pauses dans ce travail). Ce choix permet aux modèles de ne pas ignorer la non réalisation des pauses, critère tout aussi important que la durée des pauses réalisées.

Les figures 5.7 et 5.8 illustrent les fonctions de prédiction associées aux modèles de prédiction de la durée des pauses en fonction de la taille du contexte. L'unité utilisée pour décrire les tailles contextuelles est le phonème. Par soucis de lisibilité, les tailles de contextes supérieures au 9ème décile de l'ensemble des tailles observées ne sont pas affichées. La figure 5.7 correspond à la durée des pauses inter-phrases et est obtenue en utilisant une largeur de bande  $h = 0.005$ . Les durées prédites à l'aide de ce modèle varient de 427 à 912ms. La durée des pauses observées aux frontières des pseudo-syntagmes est illustrée par la figure 5.8, obtenue en utilisant une largeur de bande  $h = 0.02$ . Les durées prédites à l'aide de ce modèle varient entre 93 et 347ms.

Les modèles présentés dans ci-dessus mettent en évidence l'impact de la taille du contexte gauche et droit sur la durée des pauses. Ils peuvent être utilisés tels quels pour prédire la durée des pauses en fonction de la taille du contexte. Une de leurs limitations principales vient de leur sensibilité au bruit, notamment lorsque les données sont réparties de manière parcimonieuse. Une autre limitation, liée au choix du *Kernel*, vient de ce qu'il est difficile de contraindre l'espace des solutions explorées, par exemple en imposant que la durée soit une fonction croissante de la taille du contexte. Le système de représentation des connaissances utilisé est une *boîte noire* qui rend difficile l'explication des phénomènes observés.

### 5.6.5 Modèles paramétriques

Cette partie propose une modélisation de la durée des pauses en fonction de la taille contextuelle à l'aide de modèles paramétriques. L'observation du comportement des fonctions de prédictions de durée non paramétrique (illustré par les figures 5.7 et 5.8) sert de base pour définir un total de 15 modèles. Ces modèles sont basés sur la composition de fonctions simples : addition, multiplication, minimum, exponentielle et logarithme ; fonctions qui sont soit additionnées, soit multipliées. Les modèles définis incluent des polynômes d'ordre inférieur ou égal à 3 qui peuvent prendre en compte

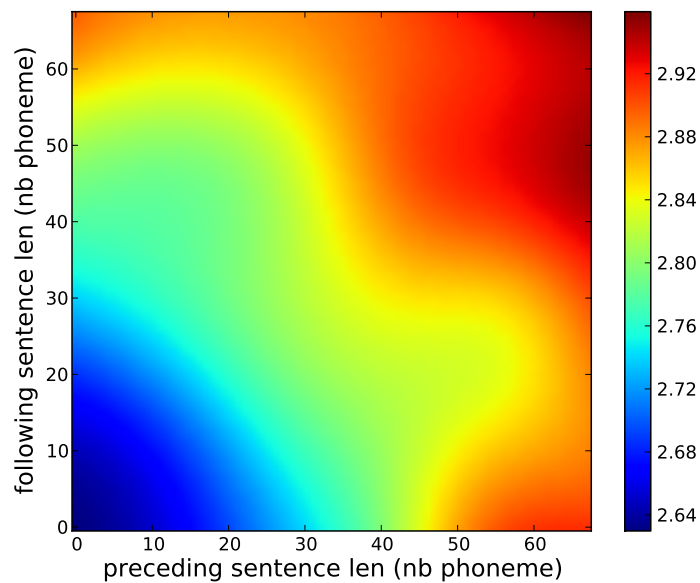


FIGURE 5.7 – Durée des pauses observées aux frontières des phrases ( $\log_{10}(ms)$ ), obtenue à l’aide d’un modèle de régression à base de *Kernel*, utilisant la taille des phrases précédentes et suivantes.

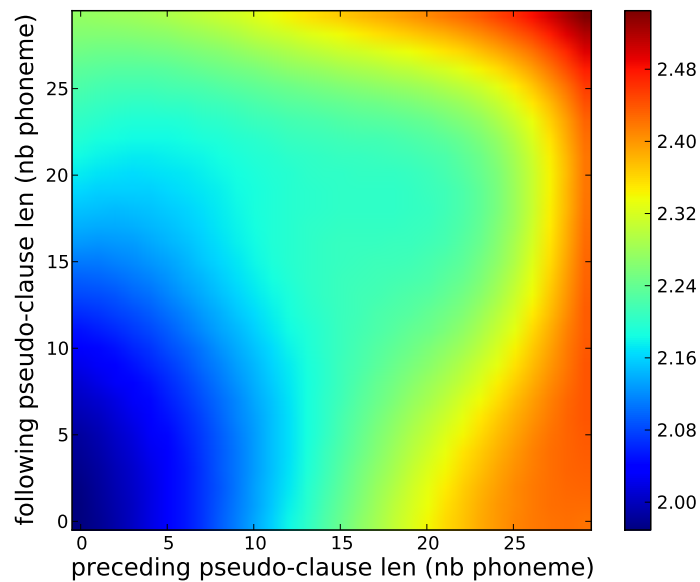


FIGURE 5.8 – Durée des pauses observées aux frontières des pseudo-syntagmes ( $\log_{10}(ms)$ ), obtenue à l’aide d’un modèle de régression à base de *Kernel*, utilisant la taille des pseudo-syntagmes précédents et suivants.

TABLE 5.7 – Modèles paramétriques utilisés pour la prédiction de la durée des pauses en fonction de la taille du contexte. Les modèles de durée ont deux arguments correspondant à la taille de l’unité précédente  $lp$ , et à la taille de l’unité suivante  $ls$ . Les paramètres des modèles  $c_i$  sont estimés sur les données à l’aide d’une procédure de régression non linéaire.

Modèle	Définition
$dur_{const}$	$c_1$
$dur_{sumlength}$	$c_1 * (lp + ls) + c_2$
$dur_{logsumlength}$	$c_1 * \log(lp + ls) + c_2$
$dur_{linpred}$	$c_1 * lp + c_2 * ls + c_3$
$dur_{loglinpred}$	$dur_{linpred}(\log(lp), \log(ls))$
$dur_{loglinpred2}$	$dur_{loglinpred}(lp + 1, ls + 1)$
$dur_{logandlin}$	$dur_{linpred} + c_4 * \log(lp + 1) + c_5 * \log(ls + 1)$
$dur_{linpredmult}$	$dur_{linpred}(lp, ls) + c_4 * lp * ls$
$dur_{logandlinandmult}$	$dur_{logandlin} + c_6 * lp * ls$
$dur_{linpredmultlog}$	$dur_{linpredmult}(\log(lp + 1), \log(ls + 1))$
$dur_{poly2}$	$dur_{linpredmult}(lp, ls) + c_5 * lp^2 + c_6 * ls^2$
$dur_{poly2log}$	$dur_{poly2}(\log(lp + 1), \log(ls + 1))$
$dur_{poly2andlog}$	$dur_{poly2}(lp, ls) + c_7 * \log(lp + 1) + c_8 * \log(ls + 1)$
$dur_{poly2andmin}$	$dur_{poly2}(lp, ls) + c_7 * \min(lp, ls) + c_8 * \min(lp^2, ls^2)$
$dur_{poly3}$	$dur_{poly2}(lp, ls) + c_7 * lp^3 + c_8 * ls^3 + c_9 * lp^2 * ls + c_{10} * lp * ls^2$

le logarithme de la taille contextuelle. Les modèles sont utilisés pour modéliser la durée des pauses à la frontière des phrases (**FPH**) et à la frontière des pseudo-syntagmes. L’ensemble des modèles considérés est détaillé dans la table 5.7.

L’évaluation des modèles qui opèrent sur les frontières des pseudo-syntagmes est réalisée dans deux conditions distinctes. Pour la première condition (**FPS**), on utilise l’ensemble des frontières de syntagmes pour l’entraînement et l’évaluation du modèle. Pour les mêmes raisons que précédemment, les frontières non associées à une pause perceptible sont associées à une durée de 40ms, ce qui permet d’inclure le phénomène de réalisation de la pause dans le modèle. Dans la deuxième condition (**FPS\***), on exclut les frontières de syntagmes qui ne sont pas associées à des pauses réalisées. Dans un cadre de synthèse de la parole, l’utilisation d’un modèle FPS\* nécessiterait de concevoir une procédure préalable permettant de prédire si la marque de ponctuation considérée doit être associée ou non à une pause.

La procédure `nlinfit` de Matlab est utilisée pour cette tâche de régression non-linéaire. Les paramètres des modèles sont optimisés afin de satisfaire le critère de minimisation des moindres carrés. Pour éviter un sur-apprentissage, les modèles sont évalués via une procédure de validation croisée qui utilise un partitionnement en 20 sous-ensembles (*20-fold cross validation*). L’ensemble des descripteurs de longueur contextuelle disponibles sont utilisés dans le cadre de cette évaluation : le nombre de phonèmes, de caractères, de syllabes et de mots. Pour un modèle donné, les meilleurs résultats sont systématiquement obtenus en utilisant les phonèmes comme unité de mesure de la taille contextuelle.

Seules les performances des modèles qui utilisent les phonèmes comme unité de description de la

TABLE 5.8 – Erreur quadratique moyenne obtenue par les modèles paramétriques de prédiction de la durée d’une pause en fonction de la taille du contexte.

Tâche	Pire modèle	Meilleur modèle	Modèle constant
FPH	0,069	0,070	0,077
FPS	0,213	0,221	0,244
FPS*	0,071	0,073	0,077

taille contextuelle sont résumées dans la table 5.8. La mesure de performance utilisée est l’erreur quadratique moyenne obtenue par les modèles sur les sous-ensembles de test. Les performances des modèles sont comparées aux performances d’un modèle retournant une valeur constante (valeur optimisant le critère des moindres carrés) afin de quantifier la contribution des informations contextuelles dans l’explication de la durée des pauses. Les moyennes des erreurs quadratiques obtenues sur les ensembles de test, associées à leur écart-types respectifs, ne permet pas de désigner un des 16 modèles comme plus approprié pour la tâche de prédiction de la durée des pauses. En conséquence, seules les moyennes des erreurs quadratiques qui correspondent aux pires et aux meilleurs modèles sont décrits dans la table. La différence de moyenne entre les pires et les meilleurs modèles est systématiquement plus petite que la différence observée avec le modèle constant minimisant le critère des moindres carrés. Ceci montre l’utilité des informations de longueur contextuelle pour la prédiction des pauses. Les prédictions du modèle opérant sur les frontières de phrases (FPH) sont plus proches des données que celles des modèles opérant sur les frontières de pseudo-syntagmes (FPS et FPS\*). L’erreur quadratique moyenne obtenue par le modèle prenant en compte les phénomènes de non réalisation de pause (FPS) est trois fois supérieure à l’erreur observée en excluant les pauses non réalisées des observations. Ce résultat met en avant l’importance de mettre au point des procédures séparées permettant de prédire la réalisation d’une pause associée à une frontière de pseudo-syntagme.

Les propriétés des modèles obtenus ont été inspectées manuellement. Les modèles linéaires ont tendance à surestimer les durées des pauses observées pour des tailles contextuelles extrêmes (très courtes ou très longues). Inversement, les modèles logarithmiques ont tendance à sous-estimer la durée des pauses pour ces tailles extrêmes. Les modèles polynomiaux ont tendance à sur-apprendre les données. Sachant que les différents modèles paramétriques obtenus ont des performances équivalentes, seuls les paramètres de la modélisation la moins complexe, basée sur un modèle linéaire, sont décrites ci-dessous. Les équations 5.3, 5.4 et 5.5 décrivent les paramètres des modèles prédisant la durée des pauses inter-phrases ( $dur_{FPH}$  et inter-pseudo-syntagmes en distinguant les cas où les phénomènes de non réalisation des pauses sont pris en compte ( $dur_{FPS}$ ) ou non ( $dur_{FPS*}$ ). La taille contextuelle est exprimée en nombre de phonèmes. Les variables  $lp$  (respectivement  $ls$ ) correspondent à la taille de l’unité précédente (ou suivante).



$$dur_{FPH}(lp, ls) = 0,0026 lp + 0,0017 ls + 2,6696 \quad (5.3)$$

$$dur_{FPS}(lp, ls) = 0,0068 lp + 0,0129 ls + 1,9521 \quad (5.4)$$

$$dur_{FPS*}(lp, ls) = 0,0031 lp + 0,0038 ls + 2,4337 \quad (5.5)$$

L'analyse des paramètres des modèles tend à montrer que la durée des pauses associées aux frontières de phrases (équation 5.3) est davantage influencée par la taille de la phrase précédant la pause, ce qui est conforme aux résultats de Zvonik (2004). À ma connaissance, la plus grande importance de la taille du pseudo-syntagme suivant une marque de ponctuation pour la prédiction de la durée de la pause, décrit dans les équations 5.4 et 5.5, n'a pas été mentionné dans la littérature. L'analyse des autres modèles (logarithmiques et polynomiaux) permet d'aboutir aux mêmes conclusions.

### 5.6.6 Impact des tours de parole sur la durée des pauses

On se propose ici de quantifier l'impact des informations correspondant aux tours de parole (cf. section 2.4.3) sur la durée des pauses. Comme cela a été montré dans les sections précédentes, la taille du contexte a une influence sur la durée des pauses qui doit être prise en compte pour mesurer l'impact des autres phénomènes liés aux propriétés linguistiques du texte. Afin de réduire ce biais, les modèles linéaires décrits précédemment (équations 5.3 et 5.5) sont utilisés pour normaliser la durée des pauses observées. Le processus de normalisation consiste donc à convertir les durées des pauses dans le domaine logarithmique, puis à les diviser par la durée estimée par les modèles de prédiction linéaire pour une taille contextuelle donnée. Ce processus est illustré par l'équation 5.6 dans le cas des pauses inter-phrases,

$$Ndur_{FPH}(d, lp, ls) = \frac{\log_{10}(\max(d, 40))}{dur_{FPH}(lp, ls)} \quad (5.6)$$

où l'argument  $d$  correspond à la durée de la pause observée dans le corpus, exprimée en millisecondes,  $dur_{FPH}(lp, ls)$  est la fonction qui prédit le  $\log_{10}$  de la durée des pauses définie par l'équation 5.3.

Les tables 5.9 et 5.10 décrivent les valeurs médianes des pauses normalisées, regroupées en catégories liées aux tours de parole. Ces catégories sont définies pour distinguer les pauses observées à l'intérieur du tour de parole du narrateur, d'un personnage, ou à une frontière de tour de parole (transition narrateur-personnage, personnage-narrateur, personnage-personnage). Les règles permettant de prédire la durée des pauses en fonction de la taille du contexte et des informations liées aux tours de parole consistent à satisfaire la condition décrite par l'équation 5.7, telle que la durée prédite, normalisée par la taille de son contexte, soit égale à la valeur normalisée médiane

TABLE 5.9 – Influence des tours de parole sur la durée des pauses inter-phrase. La durée est pauses en fonction des transitions de tours de parole est définie par  $dur_{FPH}(lp, ls)^{DTM}$ , avec  $dur_{FPH}$  définie par l'équation 5.3, et DTM la durée normalisée médiane des pauses pour ce type de transition. La dernière colonne de la table renseigne sur l'impact de la présence d'un tour de parole sur la durée d'une pause inter-phrase, avec une taille contextuelle moyenne, telle que  $dur_{FPH}(lp, ls) = 713ms$

Contexte	Nb. observations	durée normalisée médiane (DTM)	Impact sur la durée moyenne
Transition narrateur-personnage	114	1,019	808
Transition personnage-personnage différent	70	1,036	903
Même personnage	130	0,983	638
Transition personnage-narrateur	21	1,048	977
Narrateur	221	0,996	694

de la catégorie de pause considérée.

$$MedNormDurTp = \frac{\log_{10}(predTailleTp)}{\log_{10}(predTaille)} \quad (5.7)$$

où  $predTailleTp$  est la durée à prédire exprimée en millisecondes,  $predTaille$  la durée prédite par le modèle n'utilisant que la taille contextuelle (équations 5.3 ou 5.5), et  $MedNormDurTp$  est à la valeur normalisée médiane observée pour la catégorie définie par des tours de parole. La résolution de cette condition est illustrée par l'équation 5.8 :

$$predTailleTp = predTaille^{MedNormDur} \quad (5.8)$$

En conséquence, les règles de prédiction consistent à retourner la durée prédite à l'aide de modèles basés sur la taille du contexte, à la puissance correspondant à la durée médiane normalisée de la catégorie. L'impact sur la durée moyenne rapporté dans les tables correspond à l'application de l'équation 5.8 dans le cas où  $predTaille$  est égal à la durée moyenne de pause observée.

Qu'il s'agisse des pauses associées aux frontières de phrases ou de syntagmes, les durées les plus courtes sont observées à l'intérieur des tours de parole des personnages. Les frontières de phrase associées à des transitions de tour de parole ont des durées plus longues que celles observées à l'intérieur du tour de parole du narrateur. Dans le cas des incises observées aux frontières de pseudo-syntagmes, on observe des pauses courtes lors des transitions personnage-narrateur, et des pauses plus longues lorsque le personnage reprend la parole (transition narrateur-personnage).

TABLE 5.10 – Influence des tours de parole sur la durée des pauses inter-pseudo-syntagme. La durée des pauses en fonction des transitions de tours de parole est définie par  $dur_{FPS}(lp, ls)^{DTM}$ , avec  $dur_{FPS}$  définie par l'équation 5.4, et DTM la durée normalisée médiane des pauses pour ce type de transition. La dernière colonne de la table renseigne sur l'impact de la présence d'un tour de parole sur la durée d'une pause inter-pseudo-syntagme, avec une taille contextuelle moyenne, telle que  $dur_{FPS}(lp, ls) = 244ms$

Contexte	Nb. observations	durée normalisée médiane (DTM)	Impact sur la durée moyenne
Transition narrateur-personnage	32	1,066	351
même personnage	247	0,950	186
Transition personnage-narrateur	125	0,962	198
Narrateur	346	1,012	260

## 5.7 Conclusion partielle

L'analyse des variations prosodique du corpus de contes lus a montré la grande variabilité des productions du conteur. Une telle variabilité dépasse grandement celle observée dans d'autres styles de parole, même si elle montre des points communs avec le discours politique dans sa gestion des pauses. Cette variabilité est liée à de nombreux paramètres, dont un certain nombre ont montré des effets robustes et dont de nombreux autres restent à évaluer.

En premier lieu, la structure du conte se reflète dans la production du conteur. Le titre et les premiers épisodes servent à attirer l'attention de l'auditoire et montrent de nombreuses variations mélodiques, ainsi qu'une intensité moyenne assez élevée. L'élément déclencheur, qui apporte un suspens au récit, est réalisé avec une intensité moyenne clairement réduite, ce qui confirme les observations faites par [Theune et al. \(2006\)](#). De même, l'épilogue montre une baisse des variations prosodiques qui, à la différence de l'élément déclencheur, ne sert pas un suspens, mais plutôt un calme retour à la réalité, hors du conte.

L'incarnation par le même locuteur de différents personnages donne lieu à une grande variabilité prosodique. Cette variabilité est si grande qu'elle atteint parfois les limites des corrélats acoustiques classiques de la prosodie. Ainsi, l'indicateur de la hauteur perçue qu'est la fréquence fondamentale ne sert plus à grand chose pour certains personnages pour lesquels le conteur utilise une voix en grande partie dévoisée. Il est malheureusement encore difficile de caractériser acoustiquement de manière robuste les variations des paramètres de la source. On se référera utilement à [d'Alessandro \(2006\)](#) pour une description des paramètres utilisables. Il n'a donc pas été tenté d'aller plus loin dans la caractérisation de la qualité vocale car ce genre de paramètres ne sont pas disponibles pour guider le synthétiseur.

Parmi les variations observées, l'utilisation du *frequency code* ([Ohala, 1984](#)) pour informer sur la dominance relative des personnages est notable. Un tel code peut être très utile en synthèse (la F0 moyenne est un paramètre facile à contrôler), à condition de savoir déterminer ce paramètre de pouvoir.

Enfin, la gestion expressive des pauses constitue l'observation la plus aboutie de cette analyse. La durée des pauses dépend de nombreux paramètres, clairement au delà du niveau de la phrase, et l'analyse de la structure des contes nous a permis de proposer des modèles prenant en compte deux aspects de cette variation. Tout d'abord la taille du contexte (contexte syntagmatique, phrastique ou de paragraphe) permet d'estimer la durée de la pause. Cette estimation est raffinée par la gestion des tours de parole et des incises, qui viennent fournir des contraintes supplémentaires.



## Chapitre 6

# Un prototype de synthèse de parole expressive

### Sommaire

---

<b>6.1</b>	<b>Résumé du chapitre . . . . .</b>	<b>102</b>
<b>6.2</b>	<b>Architecture du système de synthèse de parole expressive . . . . .</b>	<b>102</b>
<b>6.3</b>	<b>Caractéristiques du synthétiseur de parole utilisé . . . . .</b>	<b>103</b>
6.3.1	Instructions prises en charge . . . . .	103
6.3.2	Voix de synthèse . . . . .	105
6.3.3	Outil d'édition manuelle des paramètres expressifs utilisés pour la synthèse	106
<b>6.4</b>	<b>Module d'annotation prosodique automatique . . . . .</b>	<b>107</b>
<b>6.5</b>	<b>Méthodologie d'évaluation du prototype de synthèse . . . . .</b>	<b>110</b>
<b>6.6</b>	<b>Analyse des résultats du test perceptif . . . . .</b>	<b>113</b>
6.6.1	Évaluations réalisées par les enfants . . . . .	113
6.6.2	Évaluations réalisées par les adultes . . . . .	115

---

## 6.1 Résumé du chapitre

Ce chapitre décrit le prototype de synthèse de parole expressive réalisé dans le cadre de cette thèse. Le prototype pilote le système de synthèse par sélection d'unités variables Acapela. Les instructions retenues pour moduler automatiquement la prosodie de la parole synthétisée consistent à contrôler le volume, la durée des pauses et à sélectionner la base d'unités (et donc un type de voix) la plus appropriée parmi cinq bases provenant d'un même locuteur réalisées dans le cadre du projet GV-LEx.

Ce prototype est évalué dans le cadre d'un test perceptif consistant à comparer la parole de synthèse obtenue en sortie de ce processus avec des extraits de contes lus par un locuteur professionnel, la voix de synthèse neutre correspondant au fonctionnement par défaut du synthétiseur, et une voix de synthèse ayant bénéficié de corrections manuelles après la sortie de ce prototype sans restrictions sur le jeu d'instructions prosodiques disponibles. L'analyse des résultats montre que le système proposé permet d'améliorer significativement la qualité globale de la synthèse neutre.

À l'exception d'un conte, la synthèse avec contrôle manuel des paramètres prosodiques permet d'obtenir de meilleurs résultats que la synthèse à contrôle automatique. Sans surprise, l'enregistrement correspondant au locuteur professionnel reste en toute circonstances largement préféré par les auditeurs et montre l'ampleur du chemin restant à accomplir.

## 6.2 Architecture du système de synthèse de parole expressive

Le système de synthèse de parole expressive réalisé consiste à synthétiser un fichier audio à partir d'un fichier texte. Il est constitué de trois modules principaux : un module d'analyse linguistique, un module de transformation des informations linguistiques en instructions prosodiques, et un synthétiseur de parole. Le détail des traitements et des interactions entre les différents modules est illustré par le diagramme de flux de la figure 6.1.

La communication entre les différents modules se fait à l'aide de fichiers XML. Le module d'analyse linguistique génère des fichiers respectant la DTD GV-LEx définie pour l'annotation du corpus de textes de contes décrit dans le chapitre 2. Le module de synthèse d'instructions prosodiques prend en entrée une annotation linguistique respectant la DTD GV-LEx, et génère des fichiers au format XVST, pouvant être lus et édités depuis le logiciel *Virtual Story Teller* réalisé par Acapela. Le choix des formats d'échanges est justifié par un certain nombre de contraintes liées aux spécificités du projet GV-LEx. Les deux formats d'échange retenus sont pris en charge par des logiciels d'annotation manuelle spécifiques : le logiciel d'annotation linguistique manuelle GV-LEx (section 2.3.3), et *Virtual Story Teller*, le logiciel d'annotation d'instructions prosodiques adaptées au synthétiseur Acapela. Cette spécificité permet d'envisager des scénarios de pilotage manuel, semi-automatique, ou totalement automatique du synthétiseur de parole Acapela.

Différents types de scénarios d'utilisation semi-automatiques du système sont possible. Le pre-

mier scénario consiste à utiliser les annotations linguistiques manuelles du corpus décrit dans le chapitre 2 en entrée du module de synthèse d'instructions prosodiques. Cette approche permet de tester différentes stratégies de synthèse d'instructions prosodiques en utilisant des annotations linguistiques "idéales". De cette manière, il est possible de déterminer quelles informations linguistiques permettent d'obtenir les résultats de synthèse expressive les plus encourageants. Il s'agit donc d'un scénario adapté au développement du système de synthèse, mais pas à une utilisation réelle : l'annotation linguistique manuelle de textes peut se révéler plus coûteuse qu'une annotation prosodique manuelle. Les deux autres scénarios consistent à corriger, ou enrichir manuellement les annotations linguistiques ou prosodiques obtenues via des procédures automatiques. De cette manière, il est possible de tirer parti des informations pouvant être obtenues automatiquement pour réduire le coût de l'annotation prosodique d'un texte à synthétiser. Il s'agit du cadre d'utilisation le plus viable dans un contexte industriel.

Les caractéristiques du synthétiseur de parole utilisé dans le système, ainsi que le jeu d'instructions prosodiques disponibles sont détaillés dans la section 6.3.

Le module d'analyse linguistique est constitué différentes procédures liées par des relations de dépendance. La première étape consiste à segmenter le texte en unités élémentaires à l'aide du logiciel de tokenisation *tokenisation* du LIMSI (Adda et al. (1997)). Chaque unité, ou token, est associé à un certain nombre d'informations (parties du discours, lemmes, ...) détaillées la section 3.3.2. Les unités enrichies en entrée d'un système d'annotation en tours de parole, dont la conception est détaillée dans la partie 3.3. Les tours de parole et les unités enrichies sont ensuite utilisés en entrée d'un système de détection des épisodes (section 3.4) et d'un système de détection des actes de dialogue et des modes de communication (3.5).

Le module d'annotation prosodique automatique, détaillé dans la section 6.4, fait le lien entre les informations linguistiques pouvant être extraites automatiquement, et les instructions prosodiques prise en charge par le synthétiseur Acapela, afin d'améliorer l'expressivité de la parole synthétisée. Il permet de sélectionner automatiquement la base d'unités utilisées pour la synthèse, de calculer la durée des pauses, et de faire varier le volume de lecture des différentes portions de texte considérées.

## 6.3 Caractéristiques du synthétiseur de parole utilisé

Le synthétiseur de parole utilisé dans le cadre du projet GV-LEx est le système de synthèse par sélection d'unités de tailles variables commercialisé par la société Acapela.

### 6.3.1 Instructions prises en charge

Les propriétés prosodiques de la voix générée peuvent être altérées via un certain nombre d'instructions prises en charge par le synthétiseur. Il est possible de choisir une base d'unités parmi une vingtaine de bases de données vocales disponibles en langue française. Chaque base se ca-



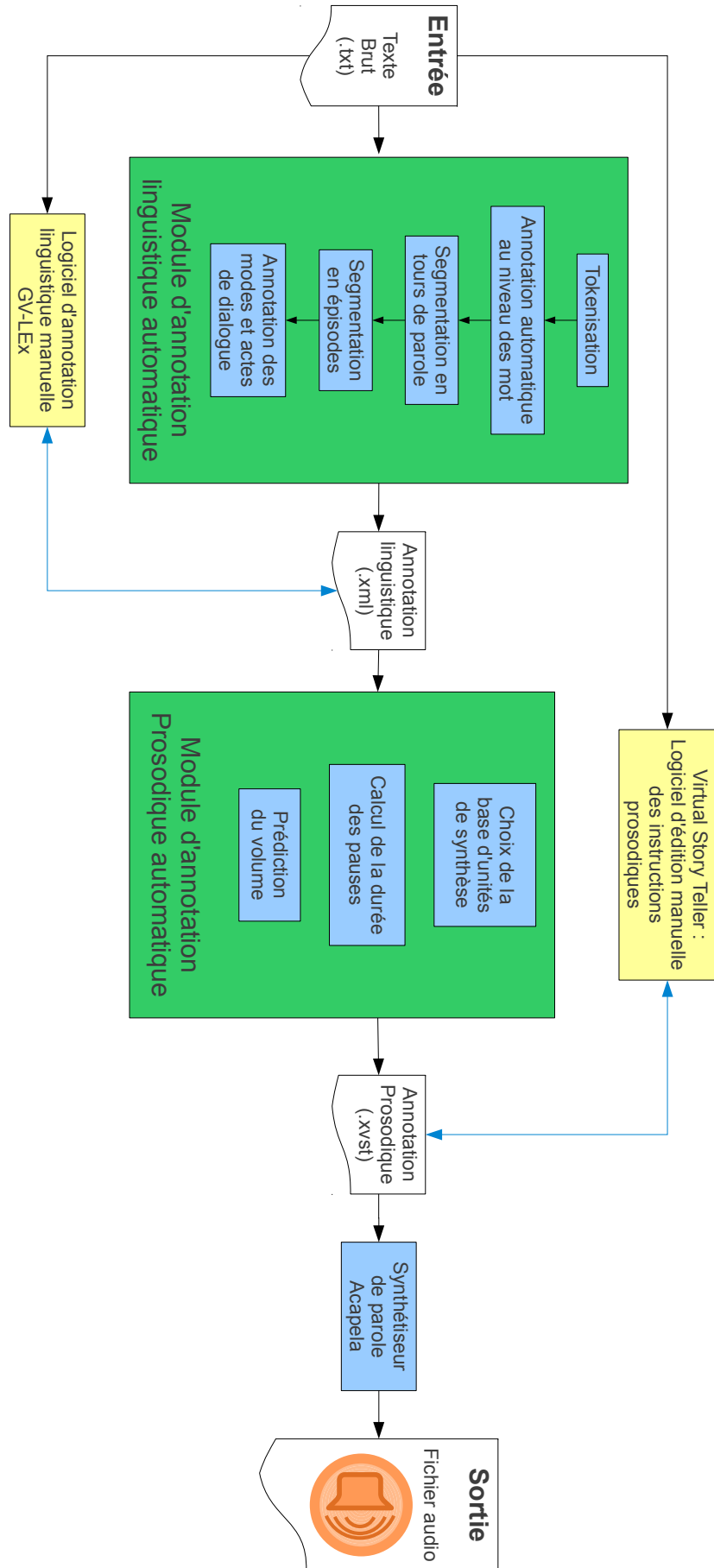


FIGURE 6.1 – Architecture du système de synthèse de parole expressive

ractérise par un certain nombre de propriétés, telles que les caractéristiques du locuteur (homme, femme, accent, etc.), ou encore le type d'expression réalisée par le locuteur lors de l'enregistrement de la base (neutre, colère, joie...). Il est également possible d'appliquer certains effets audio numériques comme des effets de transposition de la hauteur (*pitch shifting*), d'étirement temporel (*time stretching*), de modification du timbre ou du volume. De tels effets audio s'appliquent au niveau du mot : en d'autres termes, il est possible de changer la hauteur ou la durée d'un mot dans sa globalité, mais il n'est pas possible d'appliquer des modifications sur une syllabe particulière. D'autres instructions permettent de modifier le signal synthétisé : le contrôle de la réalisation et de la durée des pauses, l'ajout de *smileys vocaux* (rires, éternuements, exclamations). Il est également possible d'insérer un fichier audio externe et de le mélanger à la parole synthétisée afin de créer des ambiances sonores.

#### 6.3.2 Voix de synthèse

Dans le cadre des besoins spécifiques du projet GV-LEx, il a été décidé d'enregistrer cinq bases de synthèse différentes (l'enregistrement a été entièrement effectué par Acapela), et de les intégrer au système de synthèse. Afin de reproduire la situation d'un conteur racontant une histoire, il a été décidé que les unités utilisées pour la synthèse de parole proviendraient du même locuteur. Le choix réalisé au début du projet s'est donc porté sur les bases suivantes : une base de *voix neutre*, une de *voix souriante*, une *voix triste*, une *voix projetée* et une *voix de proximité*. Les bases de *voix de souriante*, *voix projetée* et *voix de proximité* ont été enregistrées en donnant des instructions au locuteur lui permettant de modifier explicitement la configuration de son appareil vocal. Cette particularité permet de définir des bases par rapport à des timbres de voix cibles plutôt que des affects ou des émotions, plus délicats à modéliser car associés à des dynamiques prosodiques liées au contexte. Cette stratégie contribue aussi à diminuer la part de subjectivité liée aux consignes d'enregistrement des bases. Il est en effet délicat de demander à un locuteur d'enregistrer pendant une heure une même émotion de manière réaliste et sur un contenu lexical inadapté. Dans un souci de cohérence avec l'analyse prosodique présentée dans le chapitre 4, le même locuteur professionnel a été sollicité pour enregistrer les bases de données vocales et les contes lus du corpus de parole.

La base de *voix neutre*, qui peut également être considérée comme une base de *voix narrative*, a été enregistrée en informant le locuteur que les unités collectées seraient utilisées dans un contexte de synthèse de contes pour enfants. La parole synthétisée à l'aide de cette base est a priori plus dynamique et a une plus grande variabilité mélodique, qu'une voix lue utilisée plus classiquement en synthèse. On peut comparer les motivations à l'origine de la conception de cette base aux motivations décrites dans Theune et al. (2006), qui proposent de modéliser un style de lecture adapté à la lecture d'histoires (style décrit par les auteurs comme un *global storytelling speaking style*) à l'aide de règles prosodiques adaptées à un système de synthèse de parole par concaténation de diphtonges.

La base de *voix souriante* est destinée à produire un affect positif, telle que la joie ou la gaieté, en véhiculant les paramètres acoustiques induits par l'étirement des lèvres consécutif au sourire. Pour

obtenir cette base, il a été demandé au locuteur de sourire pendant l'enregistrement des unités. Le fait prononcer des phrases en souriant permet de modifier la forme du conduit vocalique, qui impacte les propriétés formantiques du signal de parole. Cette modification produit une voix associée à des émotions positives (Tartter, 1980). On notera cependant que la transmission d'affects positifs ne se résume pas aux modifications du timbre induites par le sourire et fait intervenir d'autres modulations prosodiques (Aubergé and Cathiard, 2003) plus dynamiques. Cependant, différentes expériences ont montré cette seule modification des propriétés formantiques d'un signal de parole de synthèse, qui reproduit les propriétés du signal de parole lorsque le locuteur sourit, crée une voix associée à des percepts positifs (Mustafa et al., 2008).

La base de *voix portée*, ou *voix projetée*, a été enregistrée en donnant la consigne au locuteur de s'adresser à quelqu'un situé loin de lui. Ce type de voix correspond à des modifications de la source glottique, et en particulier à une accélération de la vitesse de fermeture de la source glottique (cf. d'Alessandro, 2006), ainsi que quelques ajustements du conduit vocal. La voix portée se caractériserait ainsi par un troisième formant relativement faible, une certaine proximité entre le troisième et le quatrième formant, une fréquence fondamentale plus élevée, et un volume perçu plus important lié à un spectre plus riche (Acker, 1987; Master et al., 2008).

La base de *voix triste*, aussi appelée *voix déprimée*, a été obtenue en demandant au locuteur d'exprimer un affect plutôt négatif. Elle se caractérise par une fréquence fondamentale relativement basse et une faible amplitude de variation mélodique.

La base de *voix de proximité* a été enregistrée en demandant au locuteur de prononcer les différentes unités comme s'il parlait à l'oreille d'un auditeur, en l'informant que la voix serait utilisée pour les situations de suspense, ou pour faire des confidences. La voix obtenue est proche du chuchotement, et se caractérise par un affaiblissement du taux de voisement.

### 6.3.3 Outil d'édition manuelle des paramètres expressifs utilisés pour la synthèse

Le logiciel Virtual Story Teller a été réalisé par Acapela pour les besoins spécifiques du projet GV-LEx. Ce logiciel permet fournir plusieurs interfaces de contrôle graphique permettant d'éditer les instructions expressives utilisées pour la synthèse, et de les convertir au format pris en charge par le synthétiseur Acapela.

Une première interface illustrée par la figure 6.2 permet d'éditer les instructions globales de synthèse consistant à structurer le texte en paragraphes. Il est possible de définir des profils de personnage caractérisés par une voix (base d'unité utilisée) ainsi qu'un volume, une vitesse, une hauteur et un timbre relatif. L'utilisateur peut ensuite structurer le texte en tours de parole, et associer un profil de personnage à chaque passage.

Une seconde interface, illustrée par la figure 6.3, permet d'éditer les paramètres expressifs locaux, tel que la durée et la localisation des pauses, l'insertion de smiley vocaux, la modulation locale des paramètres expressifs d'un profil de personnage.

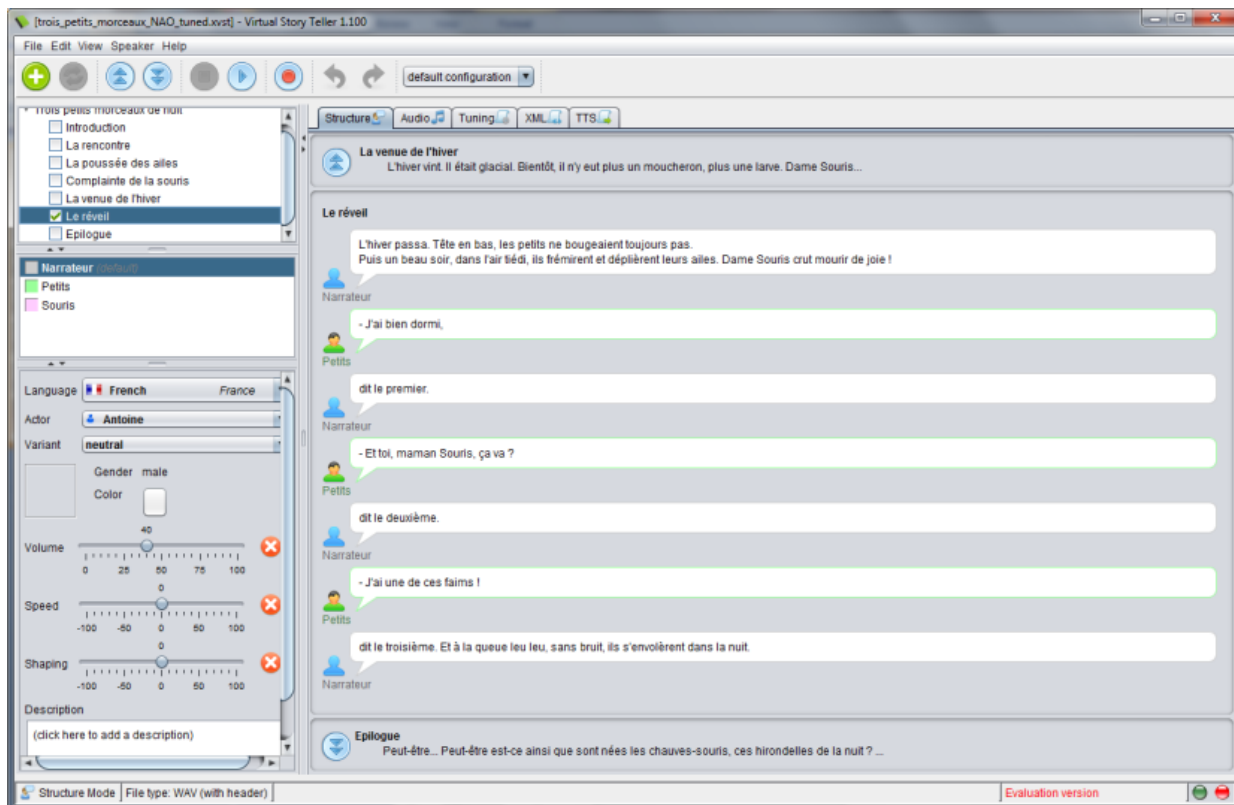


FIGURE 6.2 – Interface d’édition des propriétés expressives globales de la synthèse depuis Virtual Story Teller, logiciel d’édition manuelle d’instructions prosodiques développé par Acapela Group

Les autres interfaces de contrôle permettent d’éditer directement le fichier XML au format XVST, pouvant être lu et édité depuis Virtual Story Teller. La dernière interface permet d’éditer directement les instructions devant être réalisées par le synthétiseur.

## 6.4 Module d’annotation prosodique automatique

Le module d’annotation prosodique consiste à générer des instructions prosodiques prises en charge par le synthétiseur Acapela qui permettent de moduler la voix de synthèse, dans le but d’améliorer la lecture de contes. Pour générer ces instructions, le module utilise les informations extraites par le module d’analyse linguistique décrit dans le chapitre 3, à savoir : les frontières des épisodes, les tours de parole, les actes et les modes de dialogue, les transcriptions phonémiques et le texte brut. Le module d’annotation automatique permet de contrôler trois types de paramètres prosodiques :

- Le volume
- La base d’unités utilisées pour la synthèse
- La durée des pauses

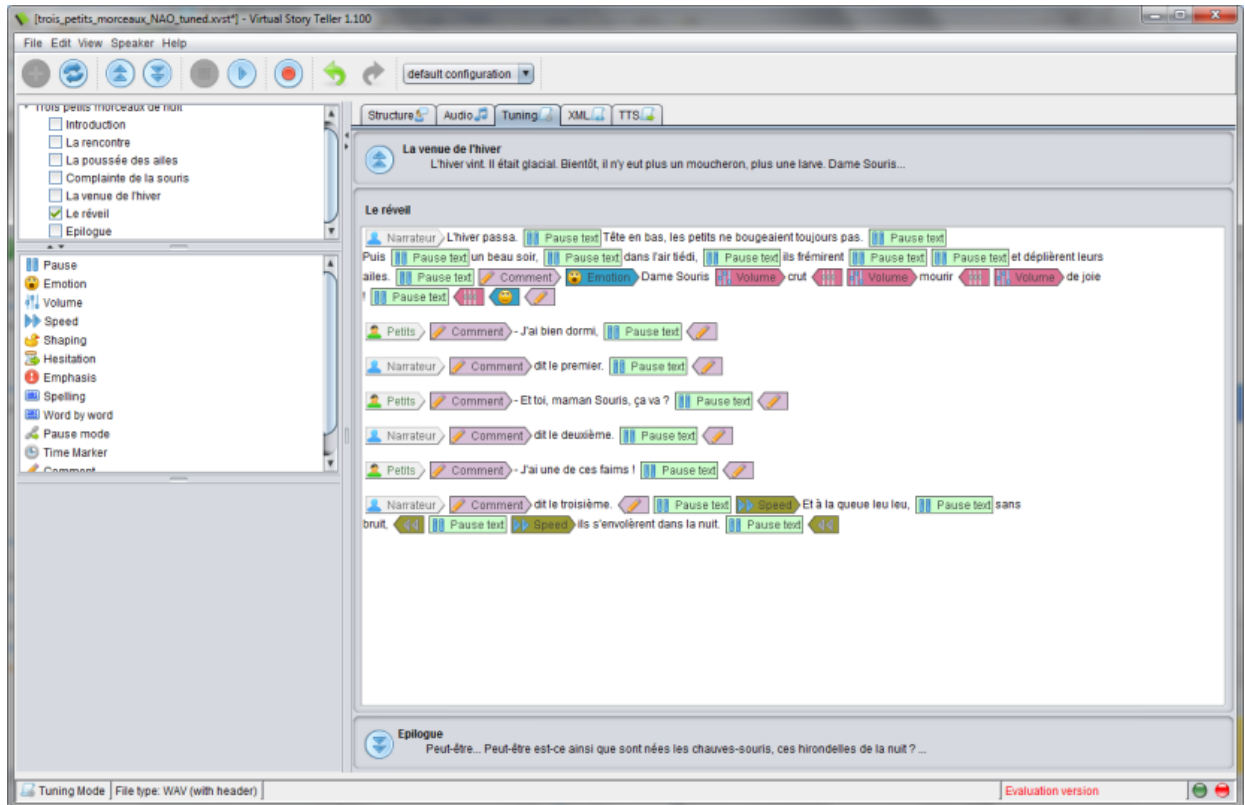


FIGURE 6.3 – Interface d'édition des propriétés expressives locales de la synthèse depuis Virtual Story Teller, logiciel d'édition manuelle d'instructions prosodiques développé par Acapela Group

tour de parole	mode et acte de dialogue	base d’unités sélectionnée
narrateur	défaut	voix neutre
narrateur	dernier épisode	voix de proximité
personnage	défaut	voix souriante
personnage	Mexclamatif + Aordre	voix projetée
personnage	Mcric	voix projetée
personnage	Mpleur	voix triste
personnage	Mrice	voix souriante
personnage	Mexclamatif	voix souriante
personnage	Anégation	voix souriante
personnage	Aordre	voix projetée
personnage	Ainterdiction	voix projetée
personnage	Aassertion	voix souriante
personnage	Aquestion	voix souriante

TABLE 6.1 – Règles de mise en relation des informations linguistiques avec le choix de la base d’unités de synthèse. Les modes sont préfixés avec la lettre M, et les actes de dialogue sont préfixés avec la lettre A. Quand seul le mode est mentionné, l’acte de dialogue est ignoré.

Les règles prosodiques inférées suite à l’analyse du corpus de parole (chapitre 4) sont appliquées directement lorsque les paramètres de contrôle du synthétiseur le permettent. Le volume associé aux tours de parole est augmenté de quatre décibels lorsqu’un personnage parle, et non plus le narrateur. Plusieurs règles permettent d’affecter la durée des pauses. La durée des pauses entre l’énoncé du titre et le reste du conte est fixée à deux secondes, la durée de la pause séparant le conte du dernier épisode (épilogue) est fixée à 1.5 secondes et les pauses associées aux autres frontières d’épisodes ont une durée fixée à 1.3 secondes. La durée des pauses inter-phrases et intra-phrase est dépendante de la longueur du contexte immédiat, exprimée en nombre de phonèmes (la longueur des pseudo-phrases, ou respectivement des pseudo-syntagmes). Les informations relatives aux tours de parole sont également utilisées pour moduler la durée des pauses, en utilisant les modèles décrits dans la section 5.6.

Les règles relatives au choix de la base d’unités de synthèse ont été déterminées empiriquement, suite à différents tests effectués depuis l’interface Virtual Story Teller décrit dans la section 6.3.3. L’ensemble des règles permettant de choisir la base d’unités utilisée pour la synthèse est illustré par la table 6.1. Les tours de parole attribués au narrateur utilisent la voix neutre, à l’exception du dernier épisode, assimilé à l’épilogue, pour lequel la voix de proximité est préférée. La base d’unités attribuée par défaut aux personnages est la voix souriante. La voix projetée est associée aux actes de langage tels que les ordres ou les interdictions mais aussi aux cris. La voix triste est associée aux pleurs. Une règle supplémentaire est utilisée dans le cas de phrases contenant des mots composés très longs (exemple 21), qui sont lus en utilisant la voix projetée.

(21) "sa-queue-est-magnifique-comme-le-soleil-quand-il-brille-sur-les-montagnes-dans-les-brouillards-du-matin" vient d’être emporté par le renard !

## 6.5 Méthodologie d'évaluation du prototype de synthèse

L'évaluation du prototype de synthèse expressive a été réalisée lors d'un test perceptif, conçu en collaboration avec Acapela. Six extraits de contes contenus dans le corpus audio GV-LEx ont été sélectionnés. Chaque extrait est associé à quatre fichiers audio, obtenus dans des conditions spécifiques.

**Locuteur Professionnel (LP)** : l'extrait correspond à l'enregistrement du locuteur professionnel réalisé lors de la constitution du corpus de parole GV-LEx (chapitre 4).

**Synthèse Neutre (SN)** : l'extrait est obtenu en utilisant la voix neutre du synthétiseur Acapela. La voix utilise le modèle prosodique par défaut du synthétiseur, traitant chaque phrase séparément.

**Synthèse Automatique (SA)** : l'extrait est obtenu en utilisant le prototype de synthèse expressive, fruit du travail de cette thèse. Les effets prosodiques sont restreints au choix d'une base d'unités parmi les 5 bases enregistrées par un même locuteur, à la modification du volume et à la gestion de la durée des pauses.

**Synthèse avec correction Manuelle (SM)** : les instructions prosodiques en sortie de la Synthèse Automatique sont éditées manuellement depuis l'interface Virtual Story Teller (développée dans ce but par Acapela, dans le cadre du projet GV-LEx). Il n'y a pas de restrictions quant au choix de la base de synthèse et les 20 bases d'Acapela disponibles pour le français peuvent être utilisées. Les smileys vocaux, les effets d'étirement temporel, de changement de hauteur et de timbre peuvent être utilisés.

Le test perceptif a été proposé aux sujets par l'intermédiaire d'une interface web. Préalablement au test, chaque participant renseignait un ensemble d'informations personnelles : nom, prénom, âge, une éventuelle habitude d'écouter de la voix de synthèse, sa langue maternelle (français, anglais ou autre) et le dispositif d'écoute utilisé (haut-parleurs ou casque). La figure 6.4 illustre les instructions données aux sujets lors de l'évaluation. Les quatre fichiers audio correspondant à un même extrait de conte sont présentés simultanément aux sujets qui devaient juger de leur qualité respective sur une échelle allant de 0 (très mauvaise qualité) à 5 (qualité optimale). Dans le cas de jeunes enfants, il a été demandé aux parents d'assister leurs enfants pendant le processus d'évaluation, pour s'assurer qu'ils comprennent bien les consignes. Tous les extraits ont été présentés comme étant synthétiques, y compris la voix naturelle. L'interface de présentation et d'évaluation des fichiers audio est illustrée par la figure 6.5. Les sujets ont la possibilité d'écouter chaque fichier audio à loisir et sont invités à laisser des commentaires en plus de leur notation. L'ordre de présentation des extraits de conte et la disposition des extraits audio au sein d'un test sont aléatoires et varient d'un sujet à l'autre. Le détail des textes utilisés lors du test est présenté dans l'annexe E. Les extraits de conte contiennent entre 42 et 111 mots, pour une taille moyenne de 77.



### GVLex Test page

---

Démarrer le test

---

Bonjour,

Ce test va vous permettre d'évaluer les qualités de conteur audio du système développé dans le cadre du projet GVLEX.

Il vous est demandé d'écouter les 4 fichiers audio et de comparer la qualité des systèmes. Pour chaque conte, merci d'écouter les 4 fichiers audio et de noter sur une échelle de 0 à 5 chaque fichier audio.

- 0 correspondant à une très mauvaise qualité,
- 5 correspondant à une qualité optimale.

Vous avez la possibilité de laisser pour chaque conte des commentaires ou jugements sur les systèmes.

Merci pour votre coopération

FIGURE 6.4 – Instructions fournies aux sujets lors de l'évaluation du système de synthèse expressive





David Doukhan  
Test en cours 1 on 6

FICHER A

▶ || ◻ 30 | 30 ?

5  
4.5  
4  
3.5  
3  
2.5  
2  
1.5  
1  
0.5  
NOTE 0

FICHER B

▶ || ◻ 18 | 28

5  
4.5  
4  
3.5  
3  
2.5  
2  
1.5  
1  
0.5  
NOTE 0

FICHER C

▶ || ◻ ?

5  
4.5  
4  
3.5  
3  
2.5  
2  
1.5  
1  
0.5  
NOTE 0

FICHER D

▶ || ◻ ?

5  
4.5  
4  
3.5  
3  
2.5  
2  
1.5  
1  
0.5  
NOTE 0

Vos commentaires :

la voix du monstre dans le fichier C est très bien

Valider mes résultats

FIGURE 6.5 – Interface d'évaluation du système de synthèse expressive

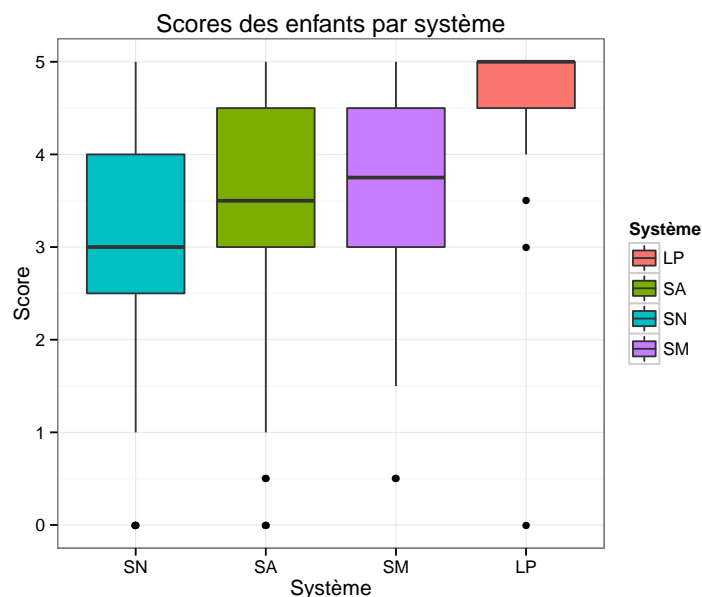


FIGURE 6.6 – Dispersion des notes données à la synthèse neutre (SN), à la synthèse automatique (SA), à la synthèse manuelle (SM) et au locuteur professionnel (LP) par les sujets de moins de 10 ans

## 6.6 Analyse des résultats du test perceptif

L'évaluation des extraits de parole a été réalisée par 77 sujets, dont dix enfants d'âge inférieur ou égal à 10 ans (et donc correspondant aux critères du projet GV-LEx). Les deux groupes de sujets sont donc étudiés séparément. Tout d'abord les dix sujets de 10 ans au plus (avec un âge moyen de 7 ans) sont analysés à l'aide d'un test sur les rangs de Kruskal-Wallis (un test statistique non-paramétrique est préféré à une ANOVA (ANalysis Of VAriance) du fait du petit nombre de sujets). Dans un second temps, les résultats obtenus par les adultes (les 67 sujets âgés de plus de 10 ans, ayant un âge moyen de 28 ans) sont analysés grâce à une analyse de variance (ANOVA).

### 6.6.1 Évaluations réalisées par les enfants

L'analyse des résultats obtenus avec les sujets les plus jeunes donne des informations sans doute moins robustes que celles fournies par les adultes, mais elle correspond au jugement du public visé par le projet GV-LEx. Du fait du petit nombre d'enfants ayant pu passer le test d'évaluation, le choix de tests statistiques non-paramétriques a été fait, de préférence à celui d'une ANOVA plus classique. Le test sur les rangs de Kruskal-Wallis montre un effet très significatif du système ayant produit les extraits sur le jugement des enfants ( $\chi^2 = 65,35$ ;  $ddl = 3$ ;  $p < 0,00001$ ). La procédure *kruskal.test()* du logiciel *R* a été utilisée. Inversement, les 6 contes qui leur sont présentés n'induisent pas de variation significative dans leurs réponses ( $\chi^2 = 3,7798$ ;  $ddl = 5$ ;  $p = 0.5815$ ). Le public visé est donc bien sensible à la stratégie de synthèse des contes qu'on leur présente.

Les distributions des résultats pour les 4 systèmes présentés sont détaillées à la figure 6.6. On

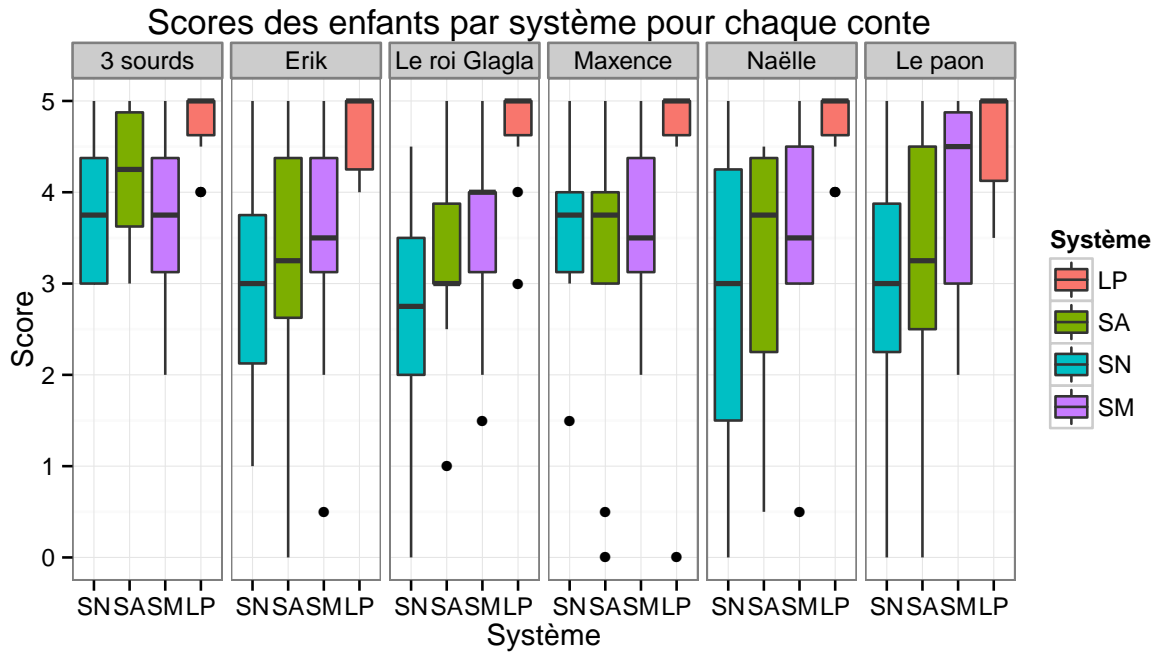


FIGURE 6.7 – Dispersion des notes données par les enfants à la synthèse neutre (SN), à la synthèse automatique (SA), à la synthèse manuelle (SM) et au locuteur professionnel (LP), pour chaque conte

Il y voit clairement que le conteur professionnel est le préféré de ce jeune public. Viennent ensuite la synthèse manuelle, suivie de la synthèse automatique et de la synthèse neutre. Les différences entre paires sont testées par un test U de Mann–Whitney. La procédure *wilcox.test()* du logiciel *R* a été utilisée. Les différences entre les scores obtenus par la synthèse manuelle et automatique, et entre la synthèse automatique et neutre ne sont donc pas significatives (les résultats pour ces deux paires sont respectivement :  $U = 1970,5$  ;  $p = 0,37$  et  $U = 1992,5$  ;  $p = 0,31$ ). Notons toutefois que la différence entre la synthèse neutre et la synthèse manuelle montre un accroissement significatif ( $U = 2181$  ;  $p < 0,05$ ).

Il y a donc bel et bien un effet induit par la stratégie de synthèse choisie. Ces différences perceptives sont plus flagrantes sur la figure 6.7, sur laquelle on peut observer de nombreuses variations induites par les différences de stratégie utilisées pour synthétiser les différents contes, par chaque système. Il est toutefois délicat de baser une analyse plus fine sur les réponses de ce jeune public, d'une part car seulement 10 sujets ont pu être interrogés mais aussi parce que ce public fait preuve d'un esprit moins analytique que le public adulte dans ses réponses, comme la figure 6.8 le montre bien. L'amplitude des notes utilisées par les enfants est clairement réduite comparée à celle des adultes. Ils fournissent une majorité de 5, note maximale, la distribution de leurs réponses allant rarement en dessous du 3 ; à l'inverse, les adultes montrent une proportion équivalente de l'usage des notes de 2 à 5. Ce manque d'esprit critique des enfants, s'il est sans doute une bonne chose pour un système de synthèse qui viserait ce public, est un défaut pour qui souhaite une évaluation constructive. La suite de l'analyse se basera donc sur les notes des 67

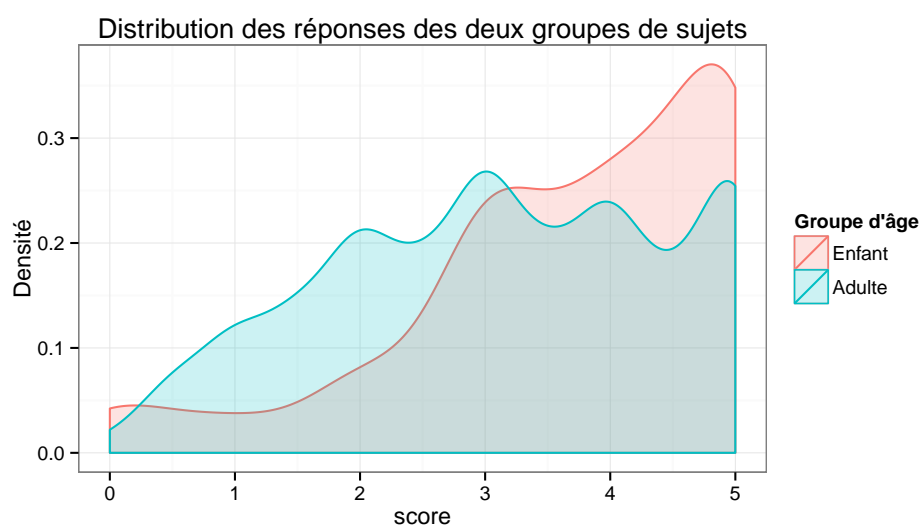


FIGURE 6.8 – Comparaison de la dispersion des notes données par les enfants et les adultes

TABLE 6.2 – Résultats de l’ANOVA appliquée aux réponses des sujets adultes

	F	ddl	ddl erreur	p	$\eta^2$ partiel
Conte	4,9250	5	1584	0,0002	0,01531
Système	458,4384	3	1584	0,0000	0,46474
Conte*Système	5,1863	15	1584	0,0000	0,04681

adultes.

### 6.6.2 Évaluations réalisées par les adultes

Une analyse de variance à deux facteurs a été menée sur les résultats obtenus par les adultes. Un facteur *Conte* (pour les 6 extraits de contes présentés aux sujets) et un facteur *Système* (pour les 4 systèmes ayant produit les stimuli) sont utilisés par l’ANOVA. La procédure *lm()* du logiciel *R* a été utilisée pour obtenir les résultats qui sont résumés dans la table 6.2. Les résultats montrent l’importance significative des deux facteurs, conte & système, sur le jugement des sujets, tout comme de leur interaction.

La colonne  $\eta^2$  partiel de la table 6.2 indique l’importance des effets de chacun des facteurs (Rietveld and Van Hout, 2005). On remarque que les différents systèmes qui ont produit les contes sont clairement à l’origine de la majeure partie des variations observées dans les résultats. L’interaction entre les facteurs *système* & *conte* explique la seconde part la plus importante de la variance (loin derrière le facteur *système*, cependant) : des particularités de synthèse propres à certains extraits peuvent expliquer des différences de résultats notables. Enfin, le facteur *conte* explique une part assez réduite de la variance des résultats.

Pour mieux comprendre ces résultats, les notes moyennes attribuées à chaque système sont

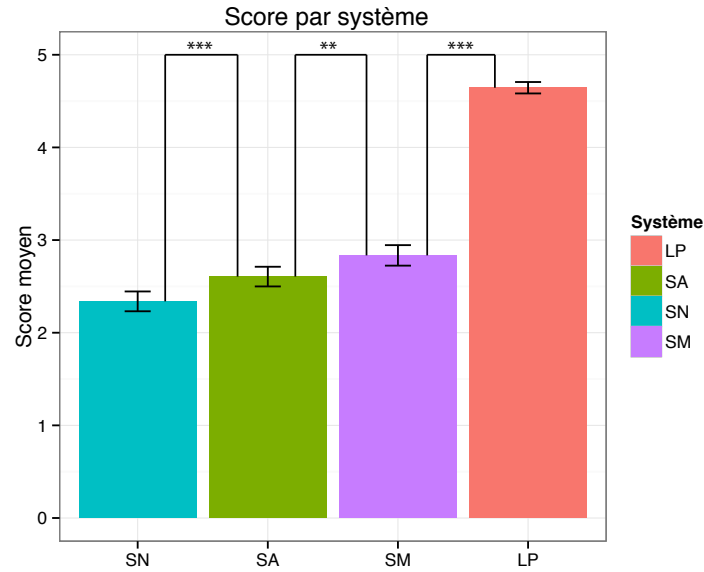


FIGURE 6.9 – Évaluation de la qualité des systèmes de synthèse. Les différences entre moyennes sont testées par un test post-hoc Tukey HSD, dont les résultats sont indiqués par les astérisques (\*\* :  $p < 0,01$  ; \*\*\* :  $p < 0,001$ ).

illustrées par la figure 6.9. Un test post-hoc (test HSD de Tukey) sur les résultats de l’ANOVA permet de tester la signification statistique des différences entre les différents niveaux des facteurs observés. La procédure *HSD.test()* du logiciel *R* a été utilisée pour effectuer ce test, sur la sortie de la procédure *lm()*. Les résultats pour le facteur *système* sont rapportés sur la figure 6.9 (les comparaisons deux-à-deux des systèmes classés par ordre de performances décroissantes sont rapportés) : les scores moyens obtenus par chaque système sont significativement différents les uns des autres, ce qui permet de classer la performance des systèmes dans un ordre strictement décroissant de performance :  $LP > SM > SA > SN$ . On observe sans surprise la grande marge qualitative entre une production humaine et une production de synthèse (une différence de l’ordre de 2 sur une échelle de 0 à 5). Les notes moyennes attribuées aux trois systèmes de synthèse sont comprises entre 2,3 et 2,8. La progression observée entre les trois systèmes de synthèse est donc significative, mais d’une amplitude toutefois beaucoup plus faible que la différence observée avec la parole naturelle. Tout cela illustre le chemin restant à parcourir pour mettre au point des systèmes de synthèse de parole à même d’égaler l’humain.

Le système obtenant les moins bons résultats correspond à la stratégie de synthèse par défaut, utilisant uniquement la voix neutre (SN) et donc non adaptée à la tâche de lecture de contes. Le système tout automatique, réalisé dans le cadre de cette thèse, fournit une performance moyenne significativement meilleure (amélioration d’une amplitude de +0,26). Ce résultat est encourageant, car il signifie que les instructions prosodiques générées automatiquement ne détériorent pas la synthèse mais au contraire contribuent à son amélioration, toujours dans le cadre de cette tâche de lecture de contes. Le système de synthèse résultant d’une correction manuelle de la sortie du système automatique fournit encore une amélioration supplémentaire (de +0,23). Le fait d’observer

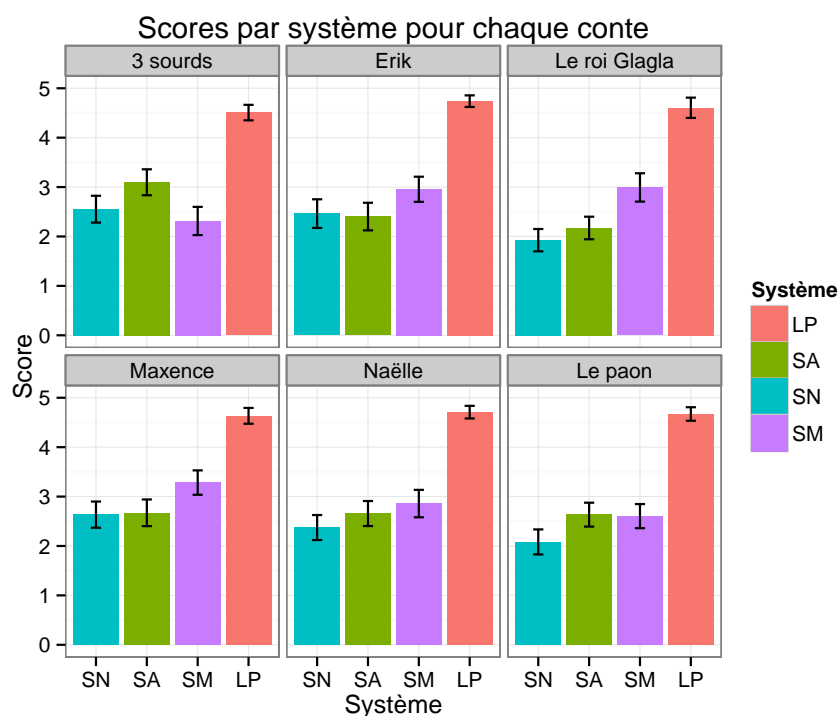


FIGURE 6.10 – Évaluation de la qualité des systèmes de synthèse par conte

une amélioration était attendu, sachant qu’il correspond à un usage expert du jeu d’instructions prosodiques, et de plus adapté précisément au contenu du conte. L’amplitude de cette amélioration n’est cependant pas plus grande que celle obtenue par le système automatique.

Une analyse plus détaillée des résultats est présentée dans la figure 6.10, qui donne les scores moyens attribués à chaque système pour chaque conte (données correspondant à l’interaction des facteurs conte & système de l’ANOVA). Les résultats du test post-hoc associé à ces données sont fournis par la table 6.3. Aucune différence significative entre les différentes stratégies de synthèse (sans prendre en compte l’humain) considérées deux à deux n’est observée pour les contes “*Naëlle va chercher du bois dans la forêt*”, “*Erik , le paysan rusé*” et “*Crétin de paon*”.

Dans le cas de “*L’histoire des trois sourds*”, les instructions prosodiques générées automatiquement permettent d’obtenir de meilleurs résultats que les instructions prosodiques manuelles (il s’agit de la seule différence significative). L’extrait sélectionné correspond à l’épilogue du conte et contient une morale. Les meilleurs résultats obtenus par le système automatique peuvent partiellement être expliqués par la règle qui associe les épilogues à l’utilisation de la voix de proximité, choix automatique qui a été modifié sur cet extrait par le système manuel et qui n’est pas celui du système neutre. Les moins bons résultats obtenus par le système de synthèse manuelle peuvent également être expliqués par l’introduction de pauses longues, qui semblent n’avoir pas été appréciées par les auditeurs.

Dans le conte “*Maxence et le monstre sous le lit*”, des différences (marginale significative entre SM & SA, significative entre SM & SN) expriment la préférence affichée pour la synthèse

TABLE 6.3 – Résultats du test HSD de Tukey mesurant la signification statistique des différences entre les différents niveaux des facteurs *système* (pour les systèmes automatiques) et *conte*

Systèmes considérés	Différence des moyennes	<i>p</i>
L’histoire des trois sourds		
SN-SA	-0,54	0,1747
SM-SA	-0,78	0,0010
SM-SN	-0,24	0,9994
Erik , le paysan rusé		
SN-SA	0,06	1,0000
SM-SA	0,55	0,1556
SM-SN	0,49	0,3550
Le roi Glagla		
SN-SA	-0,25	0,9991
SM-SA	0,82	0,0003
SM-SN	1,07	0,0000
Maxence et le monstre sous le lit		
SN-SA	-0,04	1,0000
SM-SA	0,61	0,0545
SM-SN	0,65	0,0256
Naëlle va chercher du bois dans la forêt		
SN-SA	-0,28	0,9933
SM-SA	0,20	1,0000
SM-SN	0,49	0,3868
Crétin de paon		
SN-SA	-0,55	0,1556
SM-SA	-0,03	1,0000
SM-SN	0,52	0,2422

manuelle, qui diffère notamment par l’attribution de la voix de proximité au monstre (qui dans ce cas précis se trouve être timide, chose qu’il est difficile d’appréhender de manière automatique).

Le conte du “*roi Glagla*” est celui pour lequel les instructions prosodiques manuelles ont permis d’obtenir la plus nette amélioration par rapport aux systèmes automatiques (même s’il ne s’agit pas de sa meilleure performance absolue). Dans cet extrait, un grand nombre de smileys vocaux ont été insérés manuellement. Ils correspondent aux bruits d’éternuement et à des bruits de grelottement qui avaient été enregistrés spécifiquement par Acapela dans le cadre du projet et montrent tout l’intérêt qu’il peut y avoir à utiliser des vocalisations non lexicales pour produire de la synthèse expressive.

Une première conclusion de cette comparaison détaillée des 3 systèmes de synthèse pourrait être que l’apport des règles de synthèses mises au point dans le cadre de ce travail est, certes limité,

mais jamais négatif (aucune différence négative n'est observée par rapport au système neutre de référence). Il s'agit donc d'un ensemble de règles robustes sur lesquelles il sera possible de bâtir des modifications prosodiques plus avancées. À l'inverse, on voit que les choix effectués par l'expert pour améliorer la synthèse manuellement sont, bien sûr, généralement bons, mais peuvent se révéler contre productifs. Il est donc délicat de proposer des améliorations prosodiques, et il faut baser celles-ci sur des analyses complexes des nombreux paramètres qui rentrent en ligne de compte dans ce style de lecture





# Chapitre 7

# Conclusion

## Sommaire

---

<b>7.1 Contributions de la thèse . . . . .</b>	<b>122</b>
7.1.1 Constitution, évaluation et diffusion de corpus . . . . .	122
7.1.2 Extraction automatique d'informations dans des textes fictifs . . . . .	123
7.1.3 Description des propriétés prosodiques observées dans les contes lus . . . . .	123
7.1.4 Prototype de synthèse de parole expressive . . . . .	124
<b>7.2 Perspectives de recherche . . . . .</b>	<b>124</b>
7.2.1 Extension du schéma d'annotation proposé . . . . .	124
7.2.2 Mise en application de l'ensemble des règles prosodiques formulées . . . . .	125
7.2.3 Raffinement des modèles de prédiction de la durée des pauses . . . . .	126
7.2.4 Vers une reformulation du texte . . . . .	126
<b>7.3 Publications liées à la thèse . . . . .</b>	<b>128</b>
7.3.1 Article de revue à comité de lecture . . . . .	128
7.3.2 Conférences avec actes . . . . .	128
7.3.3 Communication sans actes . . . . .	128

---

L'ensemble des travaux décrits dans ce manuscrit vise à améliorer l'expressivité des synthétiseurs de parole, en se concentrant sur le cas de la lecture d'histoires pour enfants. Le paradigme central de la thèse veut que les limitations expressives de la majorité des synthétiseurs sont liées au fait qu'ils analysent chaque phrase séparément et que la qualité globale des textes synthétisés peut être améliorée en prenant en compte les informations liées au contexte des phrases. La démarche proposée pour répondre à cette problématique a consisté à :

- définir des informations linguistiques permettant de décrire les textes *au delà du niveau de la phrase*
- créer un corpus textuel et un corpus de parole annotés selon ce formalisme
- décrire les relations observées entre les annotations linguistiques et les propriétés prosodiques estimées sur le signal de parole
- mettre au point des procédures permettant d'extraire du texte automatiquement les informations linguistiques jugées utiles pour améliorer la synthèse
- transposer les analyses prosodiques réalisées en règles permettant de contrôler les paramètres expressifs d'un synthétiseur de parole
- valider l'ensemble de la démarche via une expérience perceptive permettant de mesurer l'impact des règles définies sur la qualité globale de la parole synthétisée

## 7.1 Contributions de la thèse

### 7.1.1 Constitution, évaluation et diffusion de corpus

Un corpus textuel de 89 contes (chapitre 2) et un corpus oral de 12 contes (chapitre 4) ont été collectés, annotés, analysés et diffusés<sup>1</sup>. Les corpus sont annotés à l'aide d'informations décrivant le contenu du texte au delà de la phrase : structure narrative, référencement des mentions des personnages, citations directes attribuées et méta-informations décrivant les personnages<sup>2</sup>. Des informations décrivant les groupes lexicaux sont également utilisées et incluent des énumérations, ainsi que les entités nommées spatiales, temporelles et de personnes. Les annotations réalisées sont suffisamment génériques pour être utilisées dans d'autres contextes que la synthèse de parole expressive et répondent à un besoin croissant de données annotées sur des textes fictifs (Goh et al., 2012; Gervás, 2010; Maarouf and Villaneau, 2012). Les annotations portant sur les groupes lexicaux ne décrivent pas le texte au delà du niveau de la phrase, défini comme le paradigme majeur de la thèse. Cependant leur identification peut être considérée comme une étape préalable à l'identification de structures linguistiques de plus haut niveau (Gervás, 2010). Une analyse des accords inter-annotateurs a été effectué sur le corpus afin de décrire la reproductibilité du schéma d'annotation proposé. Les accords les plus forts sont observés pour l'identification des citations

1. Les démarches ont été effectuées auprès de l'ELRA (European Language Resources Association) pour diffuser gratuitement les corpus, ainsi que le logiciel d'annotation linguistique réalisé par Syllabs. La diffusion des corpus sera effective lors de la publication de l'article (Doukhan et al., *itre*), qui décrit exhaustivement le contenu des corpus.

2. Les méta-informations décrivant les caractéristiques des personnages n'ont été utilisées que pour annoter le corpus oral

directes, des entités de personne, ainsi que pour les chaînes de références associées aux citations directes et aux entités de personne.

### 7.1.2 Extraction automatique d'informations dans des textes fictifs

Trois systèmes permettant d'extraire des informations au delà du niveau de la phrase sont décrits dans le chapitre 3.

Un prototype de détection de tours de parole (section 3.3) consistant à identifier les zones de texte contigües correspondant à des citations directes attribuées à un unique personnage. Le prototype est basé sur une modélisation à base de CRF (Lafferty et al., 2001), ce qui le distingue des autres approches mentionnées dans la littérature, décrivant des systèmes à base de règles (Zhang et al., 2003; Mamede and Chaleira, 2004; Weiser and Watrin, 2012). Le prototype est évalué sur le corpus texte et permet d'obtenir des résultats très satisfaisants, se traduisant par une F-Mesure égale à 88,8 et un Slot Error Rate égal à 15,1.

Plusieurs méthodes de segmentation des contes en séquence d'épisodes, basées sur des extensions de l'algorithme TextTiling (Hearst, 1997) sont proposées. La première méthode utilise les tours de parole obtenus de manière automatique pour normaliser les frontières des épisodes et améliorer la qualité des segmentations prédites par TextTiling. La deuxième méthode utilise les tours de parole combinés aux références associées aux mentions des personnages obtenues manuellement, et ignore le reste du contenu des contes. Cette méthode semi-automatique permet d'obtenir les meilleures segmentations et montrent que les seules informations relatives aux mentions des personnages suffisent pour prédire les ruptures thématiques.

Le dernier système consiste à augmenter les tours de parole des personnages à l'aide d'informations relatives aux actes de dialogues (Rosset et al., 2008) et aux modes expressifs. La prédiction de ces informations est basée sur l'analyse lexicale des phrases directement adjacentes au tour de parole considéré. La qualité de ces prédictions n'est pas évaluée sur le corpus, mais déterminée empiriquement, dans le cadre de l'évaluation du prototype de synthèse expressive.

### 7.1.3 Description des propriétés prosodiques observées dans les contes lus

Des représentations graphiques sont proposées (section 4.5.4 et annexe F) pour faciliter la visualisation de la dynamique associée à plusieurs descripteurs prosodiques (hauteur, hauteur perçue, intensité et périodicité). Les propriétés prosodiques du corpus sont analysées dans le chapitre 5. Les sections 5.3 à 5.5 fournissent des descriptions basées sur l'analyse des propriétés mélodiques moyennes, en utilisant le Prosogram (Mertens, 2004) pour estimer la hauteur perçue. Les propriétés mélodiques moyennes du corpus sont comparées à différents styles de lecture et se distinguent par

une tessiture sensiblement plus vaste. Différents effets liés aux caractéristiques des personnages incarnés et au type d'épisode rencontrés sont quantifiés. L'impact de la taille du contexte et des tours de parole sur la durée des pauses observées dans le corpus est décrit dans la section 5.6. Les analyses effectuées débouchent sur la mise au point d'une méthode de normalisation des pauses en fonction de la taille du contexte, permettant de mieux observer les autres phénomènes. L'ensemble des observations est formalisé par des règles de prédiction de la durée des pauses en fonction de la taille du contexte, et des tours de parole.

#### 7.1.4 Prototype de synthèse de parole expressive

Un prototype de contrôle des paramètres prosodiques et expressifs du synthétiseur par sélection d'unités Acapela a été réalisé, permettant de contrôler le volume, la durée des pauses et le choix de la base d'unités utilisée pour la synthèse, en se basant sur une analyse du texte au delà du niveau de la phrase. Ce prototype est évalué dans le cadre d'un test perceptif consistant à comparer la parole de synthèse obtenue avec des extraits de contes lus par un locuteur professionnel, une voix de synthèse neutre et une voix de synthèse ayant bénéficié de corrections manuelles. L'analyse des résultats montre que le système proposé permet d'améliorer significativement la qualité globale de la synthèse neutre et permet de valider l'ensemble de la démarche expérimentale mise en œuvre au cours de ce manuscrit.

## 7.2 Perspectives de recherche

### 7.2.1 Extension du schéma d'annotation proposé

Les analyses prosodiques réalisées dans la section 5.5 ont mis en évidence de nombreuses relations entre les caractéristiques globales des personnages et les propriétés mélodiques moyennes observées lors de leurs prises de parole. Les méta-informations décrivant les personnages n'ont été utilisées que pour l'annotation du corpus de parole et leur utilisation pour décrire l'ensemble du corpus texte permettrait la mise au point de méthodes à même de les extraire automatiquement. Cependant, cette tâche est loin d'être triviale, car elle nécessiterait d'identifier au préalable les mentions des différents personnages, et d'attribuer les tours de parole de manière automatique.

Les paramètres expressifs du synthétiseur ont été grande partie contrôlés par les informations relatives aux actes de dialogue et aux modes expressif. Ces informations ont été obtenues à l'aide des procédures automatiques décrites dans la section 3.5, basées sur des règles syntaxiques et lexicales. Faute de corpus de référence, ces procédures n'ont pu être évaluées sur leurs capacités à améliorer la qualité globale de la parole synthétisée. Il semble donc judicieux d'annoter l'ensemble du corpus texte à l'aide de ces informations, afin de pouvoir mesurer la qualité intrinsèque des méthodes proposées et accroître leur robustesse. De telles annotations permettraient également

d’envisager d’analyser les propriétés prosodiques liées à ces informations.

Les propriétés prosodiques liées aux caractéristiques globales des personnages ont un certain nombre de limites. Il peut arriver qu’un personnage “gentil”, comme l’ours du conte Naëlle, utilise une voix très menaçante pour défendre un enfant et une voix mielleuse pour remercier un bienfaiteur. De même, les caractéristiques prosodiques communes partagées par les personnages du petit chaperon rouge et de Naëlle, les deux petites filles observées dans le corpus de parole, sont en partie liées aux similitudes de leurs personnages. Elles ne pourraient s’appliquer à des personnages tels que “Mérida”, l’héroïne du film animé “Rebelle” réalisé par les studios d’animation Pixar et sorti en 2012. Les informations liées aux modes expressifs et aux actes de dialogue semblent donc très pertinentes pour décrire le panel d’affects pouvant être exprimés par un même personnage. Il serait intéressant de comparer le pouvoir explicatif des caractéristiques globales des personnages, des actes de dialogue et des modes expressifs pour décrire les propriétés prosodiques observées sur le signal de parole.

Une autre piste d’amélioration pourrait consister à annoter le corpus de textes avec des instructions de contrôle des paramètres expressifs d’un synthétiseur cible réalisées manuellement par un ou plusieurs experts. Cette solution présente un certain nombre d’inconvénients. Le premier vient de ce que les annotations réalisées seraient dépendantes des capacités du synthétiseur à un moment donné, et devraient être modifiées en conséquence à chaque extension du jeu d’instruction supporté par le synthétiseur. Comme on a pu le voir dans le cadre de l’expérience perceptive présentée à la section 6.6.2, il peut arriver que des instructions prosodiques déterminées manuellement par un expert résultent en de moins bons résultats que de la synthèse neutre.

### 7.2.2 Mise en application de l’ensemble des règles prosodiques formulées

Les analyses prosodiques réalisées dans les parties 5.3 à 5.5 n’ont pas pu être associées à des instructions prises en charge par le synthétiseur utilisé dans le cadre du projet GV-LEx. Le synthétiseur est basé sur un moteur de sélection d’unités de tailles variables qui permet de synthétiser de la parole de haute qualité au prix d’un contrôle limité sur les propriétés mélodiques du signal synthétisé. Les instructions prosodiques supportées sont principalement basées sur la sélection de la base d’unités de synthèse à utiliser en fonction du contexte. Une analyse de la qualité vocale spécifique à chacune des cinq bases de synthèses pourrait permettre un choix automatique plus robuste, mais cela n’a pas été possible dans le cadre du projet GV-LEx. Les résultats des analyses réalisées peuvent être utilisés avec des synthétiseurs offrant plus de contrôle sur les paramètres prosodiques. Il est possible d’utiliser les résultats directement en utilisant le système décrit dans Roekhaut et al. (2010), basé sur le synthétiseur par concaténation de diphones MBROLA (Dutoit et al., 1996), ou de tirer parti des techniques de synthèse plus récentes basées sur l’utilisation de HMM (Le Maguer et al., 2012; Picart et al., 2011). Deux types de tests perceptifs peuvent être envisagés pour déterminer l’utilité des analyses réalisées. Le premier consiste à comparer, pour un

synthétiseur donné, la qualité globale de la parole synthétisée en utilisant les règles prosodiques par défaut à celle obtenue à l’aide des règles obtenues grâce aux analyses prosodiques réalisées. Un deuxième test serait de comparer la qualité globale d’un système implémentant l’ensemble des règles proposées à celle du système plus contraint basé sur le synthétiseur utilisé dans le cadre de ce travail.

### 7.2.3 Raffinement des modèles de prédiction de la durée des pauses

Les modèles proposés dans la section 5.6 permettent de prédire la durée des pauses entre les phrases et entre les pseudo-syntagmes en se basant sur la taille du contexte exprimée en nombre de phonèmes et sur les tours de parole. D’autres mesures contextuelles dont le calcul nécessite une intervention humaine ont été proposées dans la littérature pour prédire la durée des pauses, telles que la mesure de complexité syntaxique décrite dans Ferreira (1991), ou encore par la taille des phrases intonatives associée à la complexité de la structure prosodique décrites dans Krivokapić (2007). Leur utilisation dans un contexte de synthèse de parole nécessiterait de mettre au point des procédures permettant de les mesurer sans intervention humaine.

Une difficulté décrite dans Wang et al. (2008) et rencontrée par les modèles est liée aux irrégularités caractérisant le placement des pauses dans la parole expressive, et se traduit dans le corpus par un fort taux de signes de ponctuation non associés à des pauses perceptibles, ou par des pauses ne correspondant pas à des signes de ponctuation. Une étude plus poussée de ces phénomènes nécessiterait de prendre en compte les allongements des dernières syllabes (Barbosa et al., 1997), ainsi que les effets liés à la vitesse d’élocution (Grosjean and Collins, 1979).

Dans le cadre des analyses réalisées, la prédiction de la localisation des pauses rencontrées à l’intérieur des pseudo-syntagmes n’a pas été traitée, en partie parce qu’elle était prise en charge par l’analyseur syntaxique du synthétiseur Acapela. Une piste d’amélioration serait d’évaluer la pertinence des modèles basés sur une analyse syntaxique (Vannier et al., 1999) dans le cas de la parole expressive.

### 7.2.4 Vers une reformulation du texte

Dans l’introduction (section 1.1) nous avons décrit un certain nombre de techniques utilisées par les conteurs dans le cadre de leurs performances. Parmi celles-ci, on peut mentionner le travail de reformulation décrit par Muriel Bloch, consistant à essayer différentes combinaisons lexicales, jusqu’à trouver celles permettant d’être prononcées avec un rythme optimal. Dans le cadre de l’enregistrement du corpus oral (section 4.3.1), le locuteur a également été amené à effectuer quelques reformulations pour pouvoir lire le texte naturellement, notamment dans le cas de “l’histoire des trois sourds”. Lors du test perceptif décrit dans la section 6.6.2, il a également été montré que la différence d’appréciation la plus flagrante, résultant en une préférence nette pour la synthèse contrôlée manuellement, a été observée sur le conte “le roi GlaGla”, pour laquelle des

smileys vocaux correspondant à des grelottements et des éternuements, non présent explicitement dans les textes, ont été ajoutés manuellement.

Partant de ce constat, il semble évident que les capacités expressives des synthétiseurs de parole ne pourront égaler l’humain tant qu’ils ne seront pas capables d’apporter des modifications au texte devant être lu, ce qui nécessiterait de mettre au point des systèmes capables de transformer le langage écrit en langage oral. Bien qu’un tel projet nécessite un travail de très long terme, quelques pistes simples peuvent être considérées, telles que l’insertion automatique de smileys vocaux, ou encore la reformulation des incises observées par [Levin et al. \(1982\)](#) lorsque les contes sont racontés de mémoire.



## 7.3 Publications liées à la thèse

### 7.3.1 Article de revue à comité de lecture

Doukhan, D., Rosset, S., Rilliard, A., d’Alessandro, C., and Adda-Decker, M. (article soumis en mai 2013, à paraître). Gv-lex text and speech corpora: Annotated resources for extracting information and analyzing prosody of tales in french. Language Resources and Evaluation

### 7.3.2 Conférences avec actes

Doukhan, D., Rilliard, A., Rosset, S., and D’Alessandro, C. (2012a). Modelling pause duration as a function of contextual length. In InterSpeech, Portland, Oregon

Doukhan, D., Rosset, S., Rilliard, A., d’Alessandro, C., and Adda-Decker, M. (2012b). Designing french tale corpora for entertaining text to speech synthesis. In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC’12), Istanbul, Turkey. European Language Resources Association (ELRA)

Doukhan, D., Rilliard, A., Rosset, S., Adda-Decker, M., and d’Alessandro, C. (2011). Prosodic analysis of a corpus of tales. In InterSpeech, pages 3129–3132

Gelin, R., d’Alessandro, C., Le, Q., Deroo, O., Doukhan, D., Martin, J., Pelachaud, C., Rilliard, A., and Rosset, S. (2010). Towards a storytelling humanoid robot. In AAAI Fall Symposium Series on Dialog with Robots, pages 137–138

### 7.3.3 Communication sans actes

Doukhan, D. (2011). Un prototype de synthse de parole expressif. In Journées des Jeunes Chercheurs en Audition, Acoustique, Traitement du Signal Audio (JJCAAS)

Doukhan, D. (2009). Modèles de l’expressivité pour la synthèse de récits courts. In Journées des Jeunes Chercheurs en Audition, Acoustique, Traitement du Signal Audio (JJCAAS)

# Bibliographie

- Aarne, A. and Thompson, S. (1961). The types of the folktale : A classification and bibliography. Suomalainen tiedeakatemia Helsinki.
- Acker, B. (1987). Vocal tract adjustments for the projected voice. Journal of Voice, 1(1) :77–82.
- Adda, G., Adda-Decker, M., Gauvain, J., and Lamel, L. (1997). Text normalization and speech recognition in French. In EUROSPEECH, pages 2711–2714.
- Adda-Decker, M., Boula de Mareuil, P., Adda, G., and Lamel, L. (2005). Investigating syllabic structures and their variation in spontaneous French. Speech Communication, 46(2) :119–139.
- Adda-Decker, M., Gendrot, C., and Nguyen, N. (2008). Contributions du traitement automatique de la parole à l’étude des voyelles orales du français. Traitement Automatique des Langues, 49(3) :13–46.
- Adda-Decker, M. and Lamel, L. (1999). Pronunciation variants across system configuration, language and speaking style. Speech Communication, 29(2-4) :83–98.
- Adell, J., Bonafonte, A., and Escudero, D. (2005). Analysis of prosodic features towards modelling of emotional and pragmatic attributes of speech. Procesamiento de Lenguaje Natural, 35 :277–284.
- Allauzen, A. and Bonneau-Maynard, H. (2008). Training and evaluation of pos taggers on the french multitag corpus. Proceedings of the Sixth International Language Resources and Evaluation (LREC’08), Marrakech, Morocco.
- Alm, C., Roth, D., and Sproat, R. (2005). Emotions from text : machine learning for text-based emotion prediction. In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, pages 579–586. Association for Computational Linguistics.
- Alm, C. and Sproat, R. (2005). Emotional sequencing and development in fairy tales. Affective Computing and Intelligent Interaction, pages 668–674.
- Artstein, R. and Poesio, M. (2008). Inter-coder agreement for computational linguistics. Computational Linguistics, 34(4) :555–596.

- Astesano, C., Bertrand, R., Brousseau, M., Chafcouloff, M., Di Cristo, A., Ghio, A., Hirst, D., Lapierre, S., Nicolas, P., Roméas, P., et al. (1995). The PACOMUST Project, a corpus of multistyle continue speech : objectives and methodological choices. Travaux de l'institut de Phonétique d'Aix, 16 :9–38.
- Aubergé, V. and Cathiard, M. (2003). Can we hear the prosody of smile ? Speech Communication, 40(1) :87–97.
- Avanzi, M., Simon, A.-C., Goldman, J.-P., and Auchlin, A. (2010). C-prom : An annotated corpus for french prominence study. In Proceedings of Prosodic Prominence, Speech Prosody 2010 Workshop.
- Barbosa, P. and Bailly, G. (1994). Characterisation of rhythmic patterns for text-to-speech synthesis. Speech Communication, 15(1) :127–137.
- Barbosa, P. A., Bailly, G., and Barbosa, P. A. (1997). Generation of pauses within the z-score model. In Progress in Speech Synthesis, pages 365–381. Springer Verlag.
- Barras, C., Geoffrois, E., Wu, Z., and Liberman, M. (2001). Transcriber : Development and use of a tool for assisting speech corpora production. Speech Communication, 33(1-2) :5–22.
- Béchet, F. (2001). Lia phon : un systeme complet de phonétisation de textes. Traitement automatique des langues, 42(1) :47–67.
- Beeferman, D., Berger, A., and Lafferty, J. (1999). Statistical models for text segmentation. Machine learning, 34(1) :177–210.
- Bettleheim, B. (1976). The uses of enchantment. New York : Alfred A. Knopf.
- Bevacqua, E., Mancini, M., Niewiadomski, R., and Pelachaud, C. (2007). An expressive eca showing complex emotions. In Proceedings of the AISB annual convention, Newcastle, UK, pages 208–216.
- Bittar, A., Hagège, C., Moriceau, V., Tannier, X., and Teissèdre, C. (2012). Temporal annotation : A proposal for guidelines and an experiment with inter-annotator agreement. In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), Istanbul, Turkey. European Language Resources Association (ELRA).
- Blanche-Benveniste, C. (2011). Approches de la langue parlée en français. Éditions Ophrys.
- Bod, R., Fisseni, B., Kurji, A., and Löwe, B. (2012). Objectivity and reproducibility of proppian narrative annotations. In Workshop on Computational Models of Narrative.
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In Proceedings of the institute of phonetic sciences, volume 17, pages 97–110. Amsterdam.

- Boersma, P. (2002). Praat, a system for doing phonetics by computer. Glott international, 5(9/10) :341–345.
- Bozkurt, B., Dutoit, T., Prudon, R., d’Alessandro, C., and Pagel, V. (2002). Improving quality of mbrola synthesis for non-uniform units synthesis. In Speech Synthesis, 2002. Proceedings of 2002 IEEE Workshop on.
- Burkhardt, F. and Stegmann, J. (2009). Emotional speech synthesis : Applications, history and possible future. Proc. ESSV.
- Burns, P. R. (2009). Morphadorner : A java library for the morphological adornment of english language texts. In Retrieved February 12, 2013, from <http://morphadorner.northwestern.edu/>.
- Campione, E. and Véronis, J. (2004). Pauses et hésitations en français spontané. Actes des XXVè Journées d’Etude sur la Parole (JEP’2004).
- Choi, F. Y. (2000). Advances in domain independent linear text segmentation. In Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference, pages 26–33. Association for Computational Linguistics.
- Cohen, J. et al. (1960). A coefficient of agreement for nominal scales. Educational and psychological measurement, 20(1) :37–46.
- d’Alessandro, C. (2006). Voice source parameters and and prosodic analysis. In Methods in empirical prosody research, pages 63–87. Walter de Gruyter, Berlin, New York, Stefan Sudhoff, Denisa Leternovà, Roland Meyer, Sandra Pappert, Petra Augurzy, Ina Mleinek, Nicoale Richter, Johannes Schliesser edition.
- d’Alessandro, C. and Mertens, P. (1995). Automatic pitch contour stylization using a model of tonal perception. Computer Speech & Language, 9 :257–288.
- De Cheveigné, A. and Kawahara, H. (2002). Yin, a fundamental frequency estimator for speech and music. J. Acoust. Soc. Am, 111(4) :1917–1930.
- Dendien, J. and Pierrel, J.-M. (2003). Le trésor de la langue française informatisé. un exemple d’informatisation d’un dictionnaire de langue de référence. TAL. Traitement automatique des langues, 44(2) :11–37.
- Di Cristo, A. (2004). La prosodie au carrefour de la phonétique, de la phonologie et de l’articulation formes-fonctions. Travaux Interdisciplinaires du Laboratoire Parole et Langage d’Aix-en-Provence (TIPA), 23 :67–211.
- Doukhan, D. (2009). Modèles de l’expressivité pour la synthèse de récits courts. In Journées des Jeunes Chercheurs en Audition, Acoustique, Traitement du Signal Audio (JJCAAS).
- Doukhan, D. (2011). Un prototype de synthèse de parole expressif. In Journées des Jeunes Chercheurs en Audition, Acoustique, Traitement du Signal Audio (JJCAAS).

- Doukhan, D., Rilliard, A., Rosset, S., Adda-Decker, M., and d'Alessandro, C. (2011). Prosodic analysis of a corpus of tales. In InterSpeech, pages 3129–3132.
- Doukhan, D., Rilliard, A., Rosset, S., and D'Alessandro, C. (2012a). Modelling pause duration as a function of contextual length. In InterSpeech, Portland, Oregon.
- Doukhan, D., Rosset, S., Rilliard, A., d'Alessandro, C., and Adda-Decker, M. (2012b). Designing french tale corpora for entertaining text to speech synthesis. In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), Istanbul, Turkey. European Language Resources Association (ELRA).
- Doukhan, D., Rosset, S., Rilliard, A., d'Alessandro, C., and Adda-Decker, M. (article soumis en mai 2013, à paraître). Gv-lex text and speech corpora : Annotated resources for extracting information and analyzing prosody of tales in french. Language Resources and Evaluation.
- Dutoit, T., Pagel, V., Pierret, N., Bataille, F., and Van der Vrecken, O. (1996). The mbrola project : Towards a set of high quality speech synthesizers free of use for non commercial purposes. In Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on, volume 3, pages 1393–1396. IEEE.
- d'Alessandro, C., Rosset, S., and Rossi, J.-P. (1998). The pitch of short-duration fundamental frequency glissandos. The Journal of the Acoustical Society of America, 104 :2339.
- Etsion, Y. and Tsafrir, D. (2005). A short survey of commercial cluster batch schedulers. School of Computer Science and Engineering, The Hebrew University of Jerusalem, 44221 :2005–13.
- Fackrell, J., Vereecken, H., Buhmann, J., Martens, J., and Van Coile, B. (2000). Prosodic variation with text type. In Proceedings of ICSLP.
- Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. Journal of Memory and Language, 30(2) :210–233.
- Fort, K., François, C., Galibert, O., and Ghribi, M. (2012). Analyzing the Impact of Prevalence on the Evaluation of a Manual Annotation Campaign. In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), Istanbul, Turquie.
- Francisco, V., Hervás, R., Peinado, F., and Gervás, P. (2011). Emotales : creating a corpus of folk tales with emotional annotations. Language Resources and Evaluation, 45.
- Galibert, O. (2009). Approches et méthodologies pour la réponse automatique à des questions adaptées à un cadre interactif en domaine ouvert. PhD thesis, Université Paris Sud, Orsay.
- Gauvain, J., Adda, G., Adda-Decker, M., Allauzen, A., Gendner, V., Lamel, L., and Schwenk, H. (2005). Where Are We in Transcribing French Broadcast News ? In Ninth European Conference on Speech Communication and Technology. ISCA.

- Gelin, R., d'Alessandro, C., Le, Q., Deroo, O., Doukhan, D., Martin, J., Pelachaud, C., Rilliard, A., and Rosset, S. (2010). Towards a storytelling humanoid robot. In AAAI Fall Symposium Series on Dialog with Robots, pages 137–138.
- Gervás, P. (2010). Corpus annotation for narrative generation research : A wish list. In AMICUS Workshop.
- Gervás, P., Díaz-Agudo, B., Peinado, F., and Hervás, R. (2005). Story plot generation based on cbr. Knowledge-Based Systems, 18(4) :235–242.
- Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J., and Smith, N. A. (2011). Part-of-speech tagging for twitter : Annotation, features, and experiments. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Human Language Technologies : short papers-Volume 2, pages 42–47. Association for Computational Linguistics.
- Goh, H.-N., Soon, L.-K., and Haw, S.-C. (2012). Automatic identification of protagonist in fairy tales using verb. In Tan, P.-N., Chawla, S., Ho, C., and Bailey, J., editors, Advances in Knowledge Discovery and Data Mining, volume 7302 of Lecture Notes in Computer Science, pages 395–406. Springer Berlin Heidelberg.
- Gouaillier, D., Hugel, V., Blazevic, P., Kilner, C., Monceaux, J., Lafourcade, P., Marnier, B., Serre, J., and Maisonnier, B. (2009). Mechatronic design of nao humanoid. In Proceedings of the 2009 IEEE international conference on Robotics and Automation, ICRA'09, pages 2124–2129. IEEE Press.
- Grasbon, D. and Braun, N. (2001). A morphological approach to interactive storytelling. In Proc. CAST01, Living in Mixed Realities. Special issue of Netzspannung. org/journal, the Magazine for Media Production and Inter-media Research, pages 337–340. Citeseer.
- Greimas, A. (1966). Sémantique structurale : recherche et méthode. Larousse.
- Grosjean, F. and Collins, M. (1979). Breathing, pausing and reading. Phonetica, 36(2) :98–114.
- Grouin, C., Rosset, S., Zweigenbaum, P., Fort, K., Galibert, O., and Quintard, L. (2011). Proposal for an extension of traditional named entities : From guidelines to evaluation, an overview. In Proceedings of the 5th Linguistic Annotation Workshop, pages 92–100. Association for Computational Linguistics.
- Hearst, M. (1997). Texttiling : Segmenting text into multi-paragraph subtopic passages. Computational linguistics, 23(1) :33–64.
- Hendricks, W. O. (1967). On the notion 'beyond the sentence'. Linguistics, 5(37) :12–51.
- Hermes, D. J. (2006). Stylization of pitch contours. In Methods in empirical prosody research, pages 29–61. Walter de Gruyter, Berlin, New York, Stefan Sudhoff, Denisa Leternová, Roland

- Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicoale Richter, Johannes Schliesser edition.
- Hirst, D. and Di Cristo, A. (1998). A survey of intonation systems. Intonation systems : A survey of twenty languages, pages 1–44.
- Hirst, D., Di Cristo, A., and Espesser, R. (2000). Levels of representation and levels of analysis for the description of intonation systems. In Horne, M., editor, Prosody : Theory and Experiment. Studies presented to Gösta Bruce, pages 51–87. Springer.
- Hirst, D. and Espesser, R. (1993). Automatic modelling of fundamental frequency using a quadratic spline function. Travaux de l’Institut de phonétique d’Aix, 15 :71–85.
- Holt, E. (1996). Reporting on talk : The use of direct reported speech in conversation. Research on language and social interaction, 29(3) :219–245.
- Hripcsak, G. and Rothschild, A. (2005). Agreement, the f-measure, and reliability in information retrieval. Journal of the American Medical Informatics Association, 12(3) :296–298.
- Husson, F., Josse, J., and Pages, J. (2010). Principal component methods-hierarchical clustering-partitional clustering : why would we need to choose for visualizing data. Technical report, Applied Mathematics Department.
- Jun, S.-A. and Fougeron, C. (2002). Realizations of accentual phrase in french intonation. Probus, 14(1) :147–172.
- Klabbers, E. and van Santen, J. (2004). Clustering of foot-based pitch contours in expressive speech. In Proc. 5th ISCA Speech Synthesis Workshop, pages 73–78. Citeseer.
- Kohavi, R. et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In International joint Conference on artificial intelligence, volume 14, pages 1137–1145. Lawrence Erlbaum Associates Ltd.
- Krippendorff, K. (1980). Content analysis : An introduction to its methodology. Sage Publications.
- Krivokapić, J. (2007). Prosodic planning : Effects of phrasal length and complexity on pause duration. Journal of phonetics, 35(2) :162–179.
- Lafferty, J., McCallum, A., and Pereira, F. C. (2001). Conditional random fields : Probabilistic models for segmenting and labeling sequence data. In Proc. 18th International Conf. on Machine Learning, pages 282–289.
- Lamprier, S., Amghar, T., Levrat, B., and Saubion, F. (2007). On evaluation methodologies for text segmentation algorithms. In Tools with Artificial Intelligence, 2007. ICTAI 2007. 19th IEEE International Conference on, volume 2, pages 19–26. IEEE.

- Lavergne, T., Cappé, O., and Yvon, F. (2010). Practical very large scale CRFs. In Proceedings the 48th Annual Meeting of the Association for Computational Linguistics (ACL), pages 504–513. Association for Computational Linguistics.
- Le Maguer, S., Barbot, N., Boëffard, O., et al. (2012). Evaluation segmentale du système de synthèse hts pour le français (segmental evaluation of hts)[in french]. In Journées d’Etude de la Parole (JEP), conférence conjointe JEP-TALN-RECITAL.
- Leech, G. N., Garside, R., McEnery, T., et al. (1997). Corpus annotation : linguistic information from computer text corpora. Longman London, New York.
- Lendvai, P., Declerck, T., Darányi, S., Gervás, P., Hervás, R., Malec, S., and Peinado, F. (2010). Integration of linguistic markup into semantic models of folk narratives : The fairy tale use case. In LREC.
- Levin, H., Schaffer, C., and Snow, C. (1982). The prosodic and paralinguistic features of reading and telling stories. Language and speech, 25(1) :43.
- Liu, D. C. and Nocedal, J. (1989). On the limited memory bfgs method for large scale optimization. Mathematical programming, 45(1-3) :503–528.
- Loper, E. and Bird, S. (2002). Nltk : The natural language toolkit. In ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics, volume 1, pages 63–70.
- Maarouf, I. E. and Villaneau, J. (2012). A french fairy tale corpus syntactically and semantically annotated. In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC’12), Istanbul, Turkey. European Language Resources Association (ELRA).
- Makhoul, J., Kubala, F., Schwartz, R., and Weischedel, R. (1999). Performance measures for information extraction. In Proceedings of DARPA Broadcast News Workshop, pages 249–252.
- Malec, S. (2004). Proppian structural analysis and xml modeling. Proceedings of CLiP, Duisburg, Germany.
- Malec, S. (2010). Autopropp : Toward the automatic markup, classification, and annotation of russian magic tales. In First AMICUS Workshop.
- Mamede, N. and Chaleira, P. (2004). Character identification in children stories. Advances in Natural Language Processing, pages 82–90.
- Master, S., De Biase, N., Chiari, B. M., and Laukkanen, A.-M. (2008). Acoustic and perceptual analyses of brazilian male actors’ and nonactors’ voices : long-term average spectrum and the” actor’s formant”. Journal of voice : official journal of the Voice Foundation, 22(2) :146.
- Mathet, Y., Widlöcher, A., Fort, K., François, C., Galibert, O., Grouin, C., Kahn, J., Rosset, S., Zweigenbaum, P., et al. (2012). Manual corpus annotation : Giving meaning to the evaluation



- metrics. In Proceedings of the International Conference on Computational Linguistics (COLING 2012), pages 809–818.
- McCallum, A. (2002). Efficiently inducing features of conditional random fields. In Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence, pages 403–410. Morgan Kaufmann Publishers Inc.
- Mertens, P. (2004). The Prosogram : Semi-automatic transcription of prosody based on a tonal perception model. In Proceedings of Speech Prosody, pages 23–26.
- Mertens, P., Goldman, J.-P., Wehrli, E., and Gaudinat, A. (2001). La synthèse de l’intonation à partir de structures syntaxiques riches. Traitement automatique des langues, 42(1) :145–192.
- Miller, G. A. (1995). Wordnet : a lexical database for english. Communications of the ACM, 38(11) :39–41.
- Mixdorff, H. and Barbosa, P. A. (2012). Alignment of intonational events in german and brazilian portuguese—a quantitative study. In Speech Prosody 2012, pages 83–86.
- Mixdorff, H. and Fujisaki, H. (2000). A quantitative description of german prosody offering symbolic labels as a by-product. In Proceedings of ICSLP 2000, pages 98–101.
- Mustafa, A., Hashim, A.-H. A., Khalifa, O., and Ahmed, S. (2008). Toward efficient low cost highly accurate emotion speech synthesizer. IJCSNS, 8(9) :316.
- Mutlu, B., Forlizzi, J., and Hodgins, J. (2006). A storytelling robot : Modeling and evaluation of human-like gaze behavior. In Humanoid Robots, 2006 6th IEEE-RAS International Conference on, pages 518–523. IEEE.
- Nadaraya, E. (1964). On estimating regression. Teoriya Veroyatnostei i ee Primeneniya, 9(1) :157–159.
- Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of f0 of voice. Phonetica, 41(1) :1–16.
- Passonneau, R. (2004). Computing reliability for coreference annotation. In Proceedings of LREC, volume 4, pages 1503–1506.
- Patel, A., Iversen, J., and Rosenberg, J. (2006). Comparing the rhythm and melody of speech and music : The case of British English and French. The Journal of the Acoustical Society of America, 119 :3034–3047.
- Pevzner, L. and Hearst, M. (2002). A critique and improvement of an evaluation metric for text segmentation. Computational Linguistics, 28(1) :19–36.
- Picart, B., Drugman, T., and Dutoit, T. (2011). Continuous control of the degree of articulation in hmm-based speech synthesis. In INTERSPEECH, pages 1797–1800.

- Porter, M. (2001). Snowball : A language for stemming algorithms, 2001. URL :  
<http://snowball.tartarus.org/texts/introduction.html>.
- Propp, V. (1968 (orig 1928)). Morphology of the Folktale. University of Texas Press.
- Přibíl, J. and Přibílová, A. (2008). Application of expressive speech in tts system with cepstral description. In Esposito, A., Bourbakis, N. G., Avouris, N., and Hatzilygeroudis, I., editors, Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction, pages 200–212. Springer-Verlag, Berlin, Heidelberg.
- PVS, A. and Karthik, G. (2007). Part-of-speech tagging and chunking using conditional random fields and transformation based learning. Shallow Parsing for South Asian Languages, 21.
- Rabiner, L. and Juang, B.-H. (1986). An introduction to hidden markov models. ASSP Magazine, IEEE, 3(1) :4–16.
- Rajman, M., Lecomte, J., and Paroubek, P. (1997). Format de description lexicale pour le français partie 2 : Description morpho-syntaxique ref. grace gtr-3-2.1.
- Ramshaw, L. A. and Marcus, M. P. (1995). Text chunking using transformation-based learning. In Proceedings of the Third ACL Workshop on Very Large Corpora, pages 82–94. Cambridge MA, USA.
- Rietveld, T. and Van Hout, R. (2005). Statistics in language research : Analysis of variance. Walter de Gruyter.
- Roekhaut, S., Goldman, J., and Simon, A. (2010). A Model for Varying Speaking Style in TTS systems. In Fifth International Conference on Speech Prosody, Chicago, IL.
- Rosset, S., Galibert, O., Bernard, G., Bilinski, E., and Adda, G. (2009). The LIMSI multilingual, multitask QAst system. In Proc. CLEF 2008, pages 480–487, Berlin, Heidelberg. Springer-Verlag.
- Rosset, S., Tribout, D., and Lamel, L. (2008). Multi-level information and automatic dialog act detection in human–human spoken dialogs. Speech Communication, 50(1) :1–13.
- Sagot, B., Fišer, D., et al. (2008). Construction d’un wordnet libre du français à partir de ressources multilingues. In TALN 2008-Traitement Automatique des Langues Naturelles.
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In NEMLP, volume 12, pages 44–49.
- Settles, B. (2004). Biomedical named entity recognition using conditional random fields and rich feature sets. In Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications, pages 104–107. Association for Computational Linguistics.
- Shlens, J. (2014). A tutorial on principal component analysis. arXiv preprint arXiv :1404.1100.

- Sluijter, A. M. and Terken, J. M. (2009). Beyond sentence prosody : Paragraph intonation in dutch. Phonetica, 50(3) :180–188.
- Stein, A. and Schmid, H. (1995). Étiquetage morphologique de textes français avec un arbre de décisions. Traitement automatique des langues, Volume 36, Numéro 1-2 : Traitements probabilistes et corpus :23–35.
- t Hart, J. (1976). Psychoacoustic backgrounds of pitch contour stylization. IPO Annual Progress Report, 11(1) :11–19.
- Tartter, V. C. (1980). Happy talk : Perceptual and acoustic effects of smiling on speech. Perception & psychophysics, 27(1) :24–27.
- Taylor, P. (2000). Analysis and synthesis of intonation using the tilt model. The Journal of the Acoustical Society of America, 107(3) :1697–1714.
- Theune, M., Meijs, K., Heylen, D., and Ordelman, R. (2006). Generating expressive speech for storytelling applications. IEEE Transactions on Audio, Speech, and Language Processing, 14(4) :1137–1144.
- Touati, H. (2000). L’art du récit en france : état des lieux, problématique. rapport public remis au Ministère de la culture.
- Toutanova, K., Klein, D., Manning, C. D., and Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1, pages 173–180. Association for Computational Linguistics.
- Uther, H. (2004). The types of international folktales : a classification and bibliography : based on the system of Antti Aarne and Stith Thomson. Academia Scientiarum Fennica.
- Uzuner, O., Bodnari, A., Shen, S., Forbush, T., Pestian, J., and South, B. (2012). Evaluating the state of the art in coreference resolution for electronic medical records. Journal of the American Medical Informatics Association.
- van Dijk, T. (1982). Episodes as units of discourse analysis. Analyzing discourse : Text and talk, pages 177–195.
- Vandendorpe, C. (1995). Au delà de la phrase : la grammaire du texte. Pour un nouvel enseignement de la grammaire. Montréal, Éditions Logiques, pages 83–105.
- Vannier, G., Lacheret-Dujour, A., and Vergne, J. (1999). Pauses location and duration calculated with syntactic dependencies and textual considerations for tts system. ICPHS.
- Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. Information Theory, IEEE Transactions on, 13(2) :260–269.

- Wang, X., Li, A., and Yuan, C. (2008). A preliminary study on silent pauses in mandarin expressive speech. In Speech Prosody.
- Weiser, S. and Watrin, P. (2012). Extraction of unmarked quotations in newspapers. In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), Istanbul, Turkey. European Language Resources Association (ELRA).
- Zhang, J., Black, A., and Sproat, R. (2003). Identifying speakers in children's stories for speech synthesis. In Proceedings of EUROSPEECH, pages 2041–2044.
- Zhang, Z., Chapman, S., and Ciravegna, F. (2010). A methodology towards effective and efficient manual document annotation : addressing annotator discrepancy and annotation quality. In Proceedings of the 17th international conference on Knowledge engineering and management by the masses, EKAW'10, pages 301–315, Berlin, Heidelberg. Springer-Verlag.
- Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society : Series B (Statistical Methodology), 67(2) :301–320.
- Zvonik, E. (2004). Pausing and the temporal organization of phrases. An experimental study of read speech. PhD thesis, National University of Ireland.



## Annexe A

# Document Type Definition (DTD) utilisée pour l'annotation du corpus texte

```

<!ELEMENT gvlex_tale (info?,title ,exposition?,refrain?,triggering-event?,
    (scene|refrain)*,epilogue?,info?) >

<!--                                -->
<!--      marker declarations      -->
<!--                                -->

<!-- structural markers -->
<!ELEMENT title (narr|spkr)+>
<!ELEMENT triggering-event (narr|spkr)+>
<!ELEMENT exposition (narr|spkr)+>
<!ELEMENT scene (narr|spkr)+>
<!ELEMENT refrain (narr|spkr)+>
<!ELEMENT epilogue (narr|spkr)+>
<!ELEMENT info (#PCDATA|word|punct)* >

<!-- narrator speech turn, and character speech turn (direct quotation) -->
<!ELEMENT narr (pseudo_sentence)+>
<!ELEMENT spkr (pseudo_sentence)+>

<!-- restricted definition of sentence -->
<!-- should be contained in a speech turn -->
<!ELEMENT pseudo_sentence (#PCDATA|enum|person|place|time|punct|word)*>

<!-- semi-structural annotation -->
<!ELEMENT enum (#PCDATA|person|place|time|word|punct)*>

<!-- extended named entities -->
<!ELEMENT person (#PCDATA|person|place|time|word|punct)*>
<!ELEMENT time (#PCDATA|person|place|time|word|punct)*>
<!ELEMENT place (#PCDATA|person|place|time|word|punct)*>

<!-- tokens -->
<!ELEMENT word (#PCDATA) >
<!ELEMENT punct (#PCDATA) >

<!--                                -->

```

```
<!--      Attribute Declaration      -->
<!--                                     -->
```

```
<!ATTLIST scene sceneid CDATA #REQUIRED>
<!ATTLIST refrain
      occurence CDATA #REQUIRED
      refrid CDATA #REQUIRED
>
```

```
<!ATTLIST spkr persid CDATA #REQUIRED>
```

```
<!ATTLIST pseudo_sentence psid CDATA #REQUIRED>
```

```
<!ATTLIST person persid CDATA #REQUIRED>
```

```
<!ATTLIST word
      small_pos (pronoun|verb|noun|adv|adj|prep|aux|det|conj|interj)
               #REQUIRED
      large_pos CDATA #REQUIRED>
```

```
<!ATTLIST punct ptype (comma|ellipsis|rquote|lquote|semicolon|lpar|rpar
      |colon|period|dash|exclamation|interrogation) #REQUIRED>
```





## Annexe B

### Exemple de conte annoté : Le petit chaperon rouge

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE gvllex-tale SYSTEM "gvllex-1-0.dtd">

<gvllex-tale>
  <title>
    <narr>
      <pseudo-sentence psid="1">
        <person persid="1">
          <word large-pos="Da-ms-d" small-pos="det">le</word>
          <word large-pos="Afpms" small-pos="adj">petit</word>
          <word large-pos="Ncms" small-pos="noun">chaperon</word>
          <word large-pos="Afpms" small-pos="adj">rouge</word>
        </person>
      </pseudo-sentence>
    </narr>
  </title>
  <exposition>
    <narr>
      <pseudo-sentence psid="2">
        <time>
          <word large-pos="Pp3msn-" small-pos="pronoun">il</word>
          <word large-pos="Vmii3s-" small-pos="verb">était</word>
          <word large-pos="Da-fs-i" small-pos="det">une</word>
          <word large-pos="Ncfs" small-pos="noun">fois</word>
        </time>
        <person persid="1">
          <word large-pos="Da-fs-i" small-pos="det">une</word>
          <word large-pos="Afpfs" small-pos="adj">petite</word>
          <word large-pos="Ncfs" small-pos="noun">fille</word>
        </person>
        <time>
          <word large-pos="Sp" small-pos="prep">de</word>
          <word large-pos="Ncms" small-pos="noun">village</word>
        </time>
        <punct ptype="comma">,</punct>
        <word large-pos="Da-fs-d" small-pos="det">la</word>
        <word large-pos="Rgc" small-pos="adv">plus</word>
        <word large-pos="Afpfs" small-pos="adj">jolie</word>
        <word large-pos="Cs" small-pos="conj">qu'</word>
        <word large-pos="Pp3msn-" small-pos="pronoun">on</word>
        <word large-pos="Vasi3s-" small-pos="aux">eût</word>
        <word large-pos="Vmfs-sm" small-pos="verb">su</word>
        <word large-pos="Vmn—" small-pos="verb">voir</word>
        <punct ptype="semicolon">;</punct>
        <person persid="2">
          <word large-pos="Ds3fss-" small-pos="det">sa</word>
          <word large-pos="Ncfs" small-pos="noun">mère</word>
        </person>
        <word large-pos="Pp3msd-" small-pos="pronoun">en</word>
        <word large-pos="Vmii3s-" small-pos="verb">était</word>
        <word large-pos="Afpfs" small-pos="adj">folle</word>
        <punct ptype="comma">,</punct>
        <word large-pos="Cc" small-pos="conj">et</word>
        <person persid="3">
          <word large-pos="Ds3fss-" small-pos="det">sa</word>
          <word large-pos="Ncfs" small-pos="noun">mère-grand</word>
        </person>
        <word large-pos="Rgc" small-pos="adv">plus</word>
        <word large-pos="Afpfs" small-pos="adj">folle</word>
        <word large-pos="Rgp" small-pos="adv">encore</word>
        <punct ptype="period">.</punct>
      </pseudo-sentence>
      <pseudo-sentence psid="3">
        <person persid="3">
          <word large-pos="Dd-fs—" small-pos="det">cette</word>
          <word large-pos="Afpfs" small-pos="adj">bonne</word>
          <word large-pos="Ncfs" small-pos="noun">femme</word>
        </person>
        <word large-pos="Pp3msd-" small-pos="pronoun">lui</word>
        <word large-pos="Vmis3s-" small-pos="verb">fit</word>
        <word large-pos="Vmn—" small-pos="verb">faire</word>
        <word large-pos="Da-ms-i" small-pos="det">un</word>
        <word large-pos="Afpms" small-pos="adj">petit</word>
        <word large-pos="Ncms" small-pos="noun">chaperon</word>
        <word large-pos="Afpms" small-pos="adj">rouge</word>
        <punct ptype="comma">,</punct>
        <word large-pos="Pr-ms—" small-pos="pronoun">qui</word>
      </pseudo-sentence>
    </narr>
  </exposition>

```

```

    <word large_pos="Pp3msd-" small_pos="pronoun">lui</word>
  </person>
  <word large_pos="Vmip3s-" small_pos="verb">seyait</word>
  <word large_pos="Rgp" small_pos="adv">si</word>
  <word large_pos="Rgp" small_pos="adv">bien</word>
  <punct ptype="comma">,</punct>
  <word large_pos="Cs" small_pos="conj">que</word>
  <word large_pos="Rgp" small_pos="adv">partout</word>
  <word large_pos="Pp3msn-" small_pos="pronoun">on</word>
  <person persid="1">
    <word large_pos="Pp3msa-" small_pos="pronoun">l'</word>
  </person>
  <word large_pos="Vmii3s-" small_pos="verb">appelait</word>
  <word large_pos="Da-ms-d" small_pos="det">le</word>
  <word large_pos="Afpms" small_pos="adj">petit</word>
  <word large_pos="Ncms" small_pos="noun">chaperon</word>
  <word large_pos="Afpms" small_pos="adj">rouge</word>
  <punct ptype="period">.</punct>
</pseudo.sentence>
</narr>
</exposition>
<triggering-event>
  <narr>
    <pseudo.sentence psid="4">
      <place>
        <word large_pos="Da-ms-i" small_pos="det">un</word>
        <word large_pos="Ncms" small_pos="noun">jour</word>
      </place>
      <person persid="2">
        <word large_pos="Ds3fss-" small_pos="det">sa</word>
        <word large_pos="Ncfs" small_pos="noun">mère</word>
      </person>
      <punct ptype="comma">,</punct>
      <word large_pos="Vapp—" small_pos="aux">ayant</word>
      <word large_pos="Vmips-sm" small_pos="verb">cuit</word>
      <word large_pos="Cc" small_pos="conj">et</word>
      <word large_pos="Vmips-sm" small_pos="verb">fait</word>
      <word large_pos="Sp+Da-mp-d" small_pos="prep">des</word>
      <word large_pos="Ncmp" small_pos="noun">galettes</word>
      <punct ptype="comma">,</punct>
      <person persid="1">
        <word large_pos="Pp3msd-" small_pos="pronoun">lui</word>
      </person>
      <word large_pos="Vmip3s-" small_pos="verb">dit</word>
      <punct ptype="colon">:</punct>
    </pseudo.sentence>
  </narr>
  <spkr persid="2">
    <pseudo.sentence psid="5">
      <punct ptype="dash"></punct>
      <punct ptype="lquote">"</punct>
      <word large_pos="Vmip3s-" small_pos="verb">va</word>
      <word large_pos="Vmn—" small_pos="verb">voir</word>
      <word large_pos="Rxp" small_pos="adv">comment</word>
      <word large_pos="Px3ms—" small_pos="pronoun">se</word>
      <word large_pos="Vmip3s-" small_pos="verb">porte</word>
      <person persid="3">
        <word large_pos="Ds2fss-" small_pos="det">ta</word>
        <word large_pos="Ncfs" small_pos="noun">mère-grand</word>
      </person>
      <punct ptype="comma">,</punct>
      <word large_pos="Cc" small_pos="conj">car</word>
      <word large_pos="Pp3msn-" small_pos="pronoun">on</word>
      <person persid="2">
        <word large_pos="Pp1msa-" small_pos="pronoun">m'</word>
      </person>
      <word large_pos="Vaip3s-" small_pos="aux">a</word>
      <word large_pos="Vmips-sm" small_pos="verb">dit</word>
      <word large_pos="Cs" small_pos="conj">qu'</word>
      <person persid="3">
        <word large_pos="Pp3fsn-" small_pos="pronoun">elle</word>
      </person>
      <word large_pos="Vmii3s-" small_pos="verb">était</word>
      <word large_pos="Afpms" small_pos="adj">malade</word>
      <punct ptype="comma">,</punct>
      <word large_pos="Vmip3s-" small_pos="verb">porte</word>
      <person persid="3">
        <word large_pos="Pp3msd-" small_pos="pronoun">— lui</word>

```

```

</person>
<enum>
  <word large_pos="Da-fs-i" small_pos="det">une</word>
  <word large_pos="Ncfs" small_pos="noun">galette</word>
  <word large_pos="Cc" small_pos="conj">et</word>
  <word large_pos="Dd-ms—" small_pos="det">ce</word>
  <word large_pos="Afpms" small_pos="adj">petit</word>
  <word large_pos="Ncms" small_pos="noun">pot</word>
  <word large_pos="Sp" small_pos="prep">de</word>
  <word large_pos="Ncms" small_pos="noun">beurre</word>
</enum>
<punct ptype="period">.</punct>
<punct ptype="rquote">"</punct>
</pseudo_sentence>
</spkr>
</triggering-event>
<scene sceneid="1">
  <narr>
    <pseudo_sentence psid="6">
      <person persid="1">
        <word large_pos="Da-ms-d" small_pos="det">le</word>
        <word large_pos="Afpms" small_pos="adj">petit</word>
        <word large_pos="Ncms" small_pos="noun">chaperon</word>
        <word large_pos="Afpms" small_pos="adj">rouge</word>
      </person>
      <word large_pos="Vmis3s—" small_pos="verb">partit</word>
      <word large_pos="Rgp" small_pos="adv">aussitôt</word>
      <word large_pos="Sp" small_pos="prep">pour</word>
      <word large_pos="Vmn—" small_pos="verb">aller</word>
      <place>
        <word large_pos="Sp" small_pos="prep">chez</word>
      </place>
      <person persid="3">
        <word large_pos="Ds3fss—" small_pos="det">sa</word>
        <word large_pos="Ncfs" small_pos="noun">mère-grand</word>
      </person>
      </place>
      <punct ptype="comma">,</punct>
      <word large_pos="Pr-ms—" small_pos="pronoun">qui</word>
      <word large_pos="Vmii3s—" small_pos="verb">demeurait</word>
      <place>
        <word large_pos="Sp" small_pos="prep">dans</word>
        <word large_pos="Da-ms-i" small_pos="det">un</word>
        <word large_pos="Ai-ms" small_pos="adj">autre</word>
        <word large_pos="Ncms" small_pos="noun">village</word>
      </place>
      <punct ptype="period">.</punct>
    </pseudo_sentence>
  </narr>
  <narr>
    <pseudo_sentence psid="7">
      <word large_pos="Sp" small_pos="prep">en</word>
      <word large_pos="Vmpp—" small_pos="verb">passant</word>
      <place>
        <word large_pos="Sp" small_pos="prep">dans</word>
        <word large_pos="Da-ms-i" small_pos="det">un</word>
        <word large_pos="Ncms" small_pos="noun">bois</word>
      </place>
      <person persid="1">
        <word large_pos="Pp3fsn—" small_pos="pronoun">elle</word>
      </person>
      <word large_pos="Vmis3s—" small_pos="verb">rencontra</word>
      <person persid="4">
        <word large_pos="Ncms" small_pos="noun">compère</word>
        <word large_pos="Da-ms-d" small_pos="det">le</word>
        <word large_pos="Ncms" small_pos="noun">loup</word>
      </person>
      <punct ptype="comma">,</punct>
      <word large_pos="Pr-ms—" small_pos="pronoun">qui</word>
      <word large_pos="Vmis3s—" small_pos="verb">eut</word>
      <word large_pos="Rgp" small_pos="adv">bien</word>
      <word large_pos="Ncfs" small_pos="noun">envie</word>
      <word large_pos="Sp" small_pos="prep">de</word>
      <word large_pos="Pp3fsa—" small_pos="pronoun">la</word>
      <word large_pos="Vmn—" small_pos="verb">manger</word>
      <punct ptype="comma">,</punct>
      <word large_pos="Cc" small_pos="conj">mais</word>
      <person persid="4">
        <word large_pos="Pp3msn—" small_pos="pronoun">il</word>

```

```

</person>
<word large_pos="Rpn" small_pos="adv">n'</word>
<word large_pos="Vmis3s-" small_pos="verb">osa</word>
<punct ptype="comma">,</punct>
<word large_pos="Sp" small_pos="prep">à</word>
<word large_pos="Ncfs" small_pos="noun">cause</word>
<word large_pos="Sp" small_pos="prep">de</word>
<person persid="5">
  <word large_pos="Di-mp—" small_pos="det">quelques</word>
  <word large_pos="Ncmp" small_pos="noun">bûcherons</word>
</person>
<word large_pos="Pr-mp—" small_pos="pronoun">qui</word>
<word large_pos="Vmii3p—" small_pos="verb">étaient</word>
<place>
  <word large_pos="Sp" small_pos="prep">dans</word>
  <word large_pos="Da-fs-d" small_pos="det">la</word>
  <word large_pos="Ncfs" small_pos="noun">forêt</word>
</place>
<punct ptype="period">.</punct>
</pseudo.sentence>
<pseudo.sentence psid="8">
  <person persid="4">
    <word large_pos="Pp3msn—" small_pos="pronoun">il</word>
  </person>
  <person persid="1">
    <word large_pos="Pp3msd—" small_pos="pronoun">lui</word>
  </person>
  <word large_pos="Vmis3s-" small_pos="verb">demanda</word>
  <word large_pos="Pr-ms—" small_pos="pronoun">où</word>
  <person persid="1">
    <word large_pos="Pp3fsn—" small_pos="pronoun">elle</word>
  </person>
  <word large_pos="Vmii3s-" small_pos="verb">allait</word>
  <punct ptype="semicolon">;</punct>
  <person persid="1">
    <word large_pos="Da-fs-d" small_pos="det">la</word>
    <word large_pos="Afpfs" small_pos="adj">pauvre</word>
    <word large_pos="Ncfs" small_pos="noun">enfant</word>
  </person>
  <punct ptype="comma">,</punct>
  <word large_pos="Pr-ms—" small_pos="pronoun">qui</word>
  <word large_pos="Rpn" small_pos="adv">ne</word>
  <word large_pos="Vmii3s-" small_pos="verb">savait</word>
  <word large_pos="Rgn" small_pos="adv">pas</word>
  <word large_pos="Cs" small_pos="conj">qu'</word>
  <person persid="4">
    <word large_pos="Pp3msn—" small_pos="pronoun">il</word>
  </person>
  <word large_pos="Vmip3s-" small_pos="verb">est</word>
  <word large_pos="Afpms" small_pos="adj">dangereux</word>
  <word large_pos="Sp" small_pos="prep">de</word>
  <word large_pos="Px3ms—" small_pos="pronoun">s'</word>
  <word large_pos="Vmn—" small_pos="verb">arrêter</word>
  <word large_pos="Sp" small_pos="prep">à</word>
  <word large_pos="Vmn—" small_pos="verb">écouter</word>
  <person persid="4">
    <word large_pos="Da-ms-i" small_pos="det">un</word>
    <word large_pos="Ncms" small_pos="noun">loup</word>
  </person>
  <punct ptype="comma">,</punct>
  <person persid="4">
    <word large_pos="Pp3msd—" small_pos="pronoun">lui</word>
  </person>
  <word large_pos="Vmip3s-" small_pos="verb">dit</word>
  <punct ptype="colon">:</punct>
</pseudo.sentence>
</narr>
<spkr persid="1">
  <pseudo.sentence psid="9">
    <punct ptype="dash">—</punct>
    <punct ptype="lquote">"</punct>
    <person persid="1">
      <word large_pos="Pp1msn—" small_pos="pronoun">je</word>
    </person>
    <word large_pos="Vmip1s—" small_pos="verb">vais</word>
    <word large_pos="Vmn—" small_pos="verb">voir</word>
    <person persid="3">
      <word large_pos="Ds1fss—" small_pos="det">ma</word>

```

```

    <word large_pos="Ncfs" small_pos="noun">mère-grand</word>
  </person>
  <punct ptype="comma">,</punct>
  <word large_pos="Cc" small_pos="conj">et</word>
  <person persid="3">
    <word large_pos="Pp3msd-" small_pos="pronoun">lui</word>
  </person>
  <word large_pos="Vmn—" small_pos="verb">porter</word>
  <word large_pos="Da-fs-i" small_pos="det">une</word>
  <word large_pos="Ncfs" small_pos="noun">galette</word>
  <word large_pos="Sp" small_pos="prep">avec</word>
  <word large_pos="Da-ms-i" small_pos="det">un</word>
  <word large_pos="Afpms" small_pos="adj">petit</word>
  <word large_pos="Ncms" small_pos="noun">pot</word>
  <word large_pos="Sp" small_pos="prep">de</word>
  <word large_pos="Ncms" small_pos="noun">beurre</word>
  <word large_pos="Cs" small_pos="conj">que</word>
  <person persid="2">
    <word large_pos="Ds1fss-" small_pos="det">ma</word>
    <word large_pos="Ncfs" small_pos="noun">mère</word>
  </person>
  <person persid="3">
    <word large_pos="Pp3msd-" small_pos="pronoun">lui</word>
  </person>
  <word large_pos="Vmip3s-" small_pos="verb">envoie</word>
  <punct ptype="period">.</punct>
  <punct ptype="quote">"</punct>
</pseudo_sentence>
</spkr>
<spkr persid="4">
  <pseudo_sentence psid="10">
    <punct ptype="dash">—</punct>
    <punct ptype="lquote">"</punct>
    <word large_pos="Vmip3s-" small_pos="verb">demeure</word>
    <person persid="3">
      <word large_pos="Pp3fsn-" small_pos="pronoun">—t-elle</word>
    </person>
    <word large_pos="Rgp" small_pos="adv">bien</word>
    <word large_pos="Rgp" small_pos="adv">loin</word>
    <punct ptype="interrogation">?</punct>
    <punct ptype="quote">"</punct>
  </pseudo_sentence>
</spkr>
<narr>
  <pseudo_sentence psid="11">
    <person persid="1">
      <word large_pos="Pp3msd-" small_pos="pronoun">lui</word>
    </person>
    <word large_pos="Vmip3s-" small_pos="verb">dit</word>
    <person persid="4">
      <word large_pos="Da-ms-d" small_pos="det">le</word>
      <word large_pos="Ncms" small_pos="noun">loup</word>
    </person>
    <punct ptype="period">.</punct>
  </pseudo_sentence>
</narr>
<spkr persid="1">
  <pseudo_sentence psid="12">
    <punct ptype="dash">—</punct>
    <punct ptype="lquote">"</punct>
    <word large_pos="I" small_pos="interj">oh</word>
    <punct ptype="exclamation">!</punct>
    <word large_pos="Rgp" small_pos="adv">oui</word>
    <punct ptype="quote">"</punct>
    <punct ptype="comma">,</punct>
  </pseudo_sentence>
</spkr>
<narr>
  <pseudo_sentence psid="13">
    <word large_pos="Vmip3s-" small_pos="verb">dit</word>
    <person persid="1">
      <word large_pos="Da-ms-d" small_pos="det">le</word>
      <word large_pos="Afpms" small_pos="adj">petit</word>
      <word large_pos="Ncms" small_pos="noun">chaperon</word>
      <word large_pos="Afpms" small_pos="adj">rouge</word>
    </person>
    <punct ptype="comma">,</punct>
  </pseudo_sentence>

```

```

</narr>
<spkr persid="1">
  <pseudo.sentence psid="14">
    <punct ptype="lquote">"</punct>
    <word large_pos="Pd-ms—" small_pos="pronoun">c'</word>
    <word large_pos="Vmip3s—" small_pos="verb">est</word>
    <place>
      <word large_pos="Sp" small_pos="prep">par</word>
      <word large_pos="Rgp" small_pos="adv">delà</word>
      <word large_pos="Da-ms-d" small_pos="det">le</word>
      <word large_pos="Ncms" small_pos="noun">moulin</word>
    </place>
    <word large_pos="Cs" small_pos="conj">que</word>
    <word large_pos="Pp2mpn—" small_pos="pronoun">vous</word>
    <word large_pos="Vmip2p—" small_pos="verb">voyez</word>
    <place>
      <word large_pos="Rgp" small_pos="adv">tout</word>
      <word large_pos="Rgp" small_pos="adv">là-bas</word>
    </place>
    <punct ptype="comma">,</punct>
    <place>
      <word large_pos="Rgp" small_pos="adv">là-bas</word>
    </place>
    <punct ptype="comma">,</punct>
    <place>
      <word large_pos="Sp" small_pos="prep">à</word>
      <word large_pos="Da-fs-d" small_pos="det">la</word>
      <word large_pos="Ao-fs" small_pos="adj">première</word>
      <word large_pos="Ncfs" small_pos="noun">maison</word>
      <word large_pos="Sp+Da-ms-d" small_pos="prep">du</word>
      <word large_pos="Ncms" small_pos="noun">village</word>
    </place>
    <punct ptype="period">.</punct>
    <punct ptype="rquote">"</punct>
  </pseudo.sentence>
</spkr>
<spkr persid="4">
  <pseudo.sentence psid="15">
    <punct ptype="dash">"</punct>
    <punct ptype="lquote">"</punct>
    <word large_pos="I" small_pos="interj">eh</word>
    <word large_pos="Rgp" small_pos="adv">bien</word>
    <punct ptype="exclamation">!"</punct>
    <punct ptype="rquote">"</punct>
    <punct ptype="comma">,</punct>
  </pseudo.sentence>
</spkr>
<narr>
  <pseudo.sentence psid="16">
    <word large_pos="Vmip3s—" small_pos="verb">dit</word>
    <person persid="4">
      <word large_pos="Da-ms-d" small_pos="det">le</word>
      <word large_pos="Ncms" small_pos="noun">loup</word>
    </person>
    <punct ptype="comma">,</punct>
  </pseudo.sentence>
</narr>
<spkr persid="4">
  <pseudo.sentence psid="17">
    <punct ptype="lquote">"</punct>
    <person persid="4">
      <word large_pos="Pp1msn—" small_pos="pronoun">je</word>
    </person>
    <word large_pos="Vmip1s—" small_pos="verb">veux</word>
    <word large_pos="Pp3msd—" small_pos="pronoun">y</word>
    <word large_pos="Vmn—" small_pos="verb">aller</word>
    <word large_pos="Vmn—" small_pos="verb">voir</word>
    <word large_pos="Rgp" small_pos="adv">aussi</word>
    <punct ptype="semicolon">;</punct>
    <person persid="4">
      <word large_pos="Pp1msn—" small_pos="pronoun">je</word>
    </person>
    <person persid="4">
      <word large_pos="Px1ms—" small_pos="pronoun">m'</word>
    </person>
    <word large_pos="Pp3msd—" small_pos="pronoun">y</word>
    <word large_pos="Pp3msd—" small_pos="pronoun">en</word>
    <word large_pos="Vmip1s—" small_pos="verb">vais</word>
  </pseudo.sentence>
</spkr>

```



```

<word large_pos="Sp" small_pos="prep">par</word>
<word large_pos="Dd-ms—" small_pos="det">ce</word>
<word large_pos="Ncms" small_pos="noun">chemin</word>
<word large_pos="Rgp" small_pos="adv">-ci</word>
<punct ptype="comma">,</punct>
<word large_pos="Cc" small_pos="conj">et</word>
<person persid="1">
  <word large_pos="Pp2mso—" small_pos="pronoun">toi</word>
</person>
<word large_pos="Sp" small_pos="prep">par</word>
<word large_pos="Dd-ms—" small_pos="det">ce</word>
<word large_pos="Ncms" small_pos="noun">chemin</word>
<word large_pos="Rgp" small_pos="adv">-là</word>
<punct ptype="comma">,</punct>
<word large_pos="Cc" small_pos="conj">et</word>
<person persid="6">
  <word large_pos="Pp1mpn—" small_pos="pronoun">nous</word>
</person>
<word large_pos="Vmif1p—" small_pos="verb">verrons</word>
<word large_pos="Pr-mp—" small_pos="pronoun">qui</word>
<word large_pos="Rgc" small_pos="adv">plus</word>
<word large_pos="Rgp" small_pos="adv">tôt</word>
<word large_pos="Pp3msd—" small_pos="pronoun">y</word>
<word large_pos="Vmif3s—" small_pos="verb">sera</word>
<punct ptype="period">.</punct>
<punct ptype="rquote">"</punct>
</pseudo_sentence>
</spkr>
</scene>
<scene sceneid="2">
  <narr>
    <pseudo_sentence psid="18">
      <person persid="4">
        <word large_pos="Da-ms-d" small_pos="det">le</word>
        <word large_pos="Ncms" small_pos="noun">loup</word>
      </person>
      <word large_pos="Px3ms—" small_pos="pronoun">se</word>
      <word large_pos="Vmis3s—" small_pos="verb">mit</word>
      <word large_pos="Sp" small_pos="prep">à</word>
      <word large_pos="Vmn—" small_pos="verb">courir</word>
      <word large_pos="Sp" small_pos="prep">de</word>
      <word large_pos="Di-fs—" small_pos="det">toute</word>
      <word large_pos="Ds3fss—" small_pos="det">sa</word>
      <word large_pos="Ncfs" small_pos="noun">force</word>
      <word large_pos="Sp" small_pos="prep">par</word>
      <word large_pos="Da-ms-d" small_pos="det">le</word>
      <word large_pos="Ncms" small_pos="noun">chemin</word>
      <word large_pos="Pr-ms—" small_pos="pronoun">qui</word>
      <word large_pos="Vmii3s—" small_pos="verb">était</word>
      <word large_pos="Da-ms-d" small_pos="det">le</word>
      <word large_pos="Rgc" small_pos="adv">plus</word>
      <word large_pos="Afpms" small_pos="adj">court</word>
      <punct ptype="comma">,</punct>
      <word large_pos="Cc" small_pos="conj">et</word>
      <person persid="1">
        <word large_pos="Da-fs-d" small_pos="det">la</word>
        <word large_pos="Afpfs" small_pos="adj">petite</word>
        <word large_pos="Ncfs" small_pos="noun">fille</word>
      </person>
      <word large_pos="Px3fs—" small_pos="pronoun">s'</word>
      <word large_pos="Pp3msd—" small_pos="pronoun">en</word>
      <word large_pos="Vmis3s—" small_pos="verb">alla</word>
      <word large_pos="Sp" small_pos="prep">par</word>
      <word large_pos="Da-ms-d" small_pos="det">le</word>
      <word large_pos="Ncms" small_pos="noun">chemin</word>
      <word large_pos="Da-ms-d" small_pos="det">le</word>
      <word large_pos="Rgc" small_pos="adv">plus</word>
      <word large_pos="Afpms" small_pos="adj">long</word>
      <punct ptype="comma">,</punct>
      <word large_pos="Px3ms—" small_pos="pronoun">s'</word>
      <word large_pos="Vmpp—" small_pos="verb">amusant</word>
    </pseudo_sentence>
    <word large_pos="Sp" small_pos="prep">à</word>
    <word large_pos="Vmn—" small_pos="verb">cueillir</word>
    <word large_pos="Da-fp-i" small_pos="det">des</word>
    <word large_pos="Ncfp" small_pos="noun">noisettes</word>
    <punct ptype="comma">,</punct>
    <word large_pos="Sp" small_pos="prep">à</word>
  </narr>

```

```

<word large_pos="Vmn—" small_pos="verb">courir</word>
<word large_pos="Sp" small_pos="prep">après</word>
<word large_pos="Da-mp-i" small_pos="det">des</word>
<word large_pos="Ncmp" small_pos="noun">papillons</word>
<punct ptype="comma">,</punct>
<word large_pos="Cc" small_pos="conj">et</word>
<word large_pos="Sp" small_pos="prep">à</word>
<word large_pos="Vmn—" small_pos="verb">faire</word>
<word large_pos="Da-mp-i" small_pos="det">des</word>
<word large_pos="Ncmp" small_pos="noun">bouquets</word>
<word large_pos="Sp+Da-fp-d" small_pos="prep">des</word>
<word large_pos="Afpfp" small_pos="adj">petites</word>
<word large_pos="Ncfp" small_pos="noun">fleurs</word>
<word large_pos="Pr-fp—" small_pos="pronoun">qu'</word>
<person persid="1">
  <word large_pos="Pp3fsn—" small_pos="pronoun">elle</word>
</person>
<word large_pos="Vmii3s—" small_pos="verb">rencontrait</word>
</enum>
<punct ptype="period">.</punct>
</pseudo_sentence>
<pseudo_sentence psid="19">
  <person persid="4">
    <word large_pos="Da-ms-d" small_pos="det">le</word>
    <word large_pos="Ncms" small_pos="noun">loup</word>
  </person>
  <word large_pos="Rpn" small_pos="adv">ne</word>
  <word large_pos="Vais3s—" small_pos="aux">fut</word>
  <word large_pos="Rgn" small_pos="adv">pas</word>
  <word large_pos="Rgp" small_pos="adv">longtemps</word>
  <word large_pos="Sp" small_pos="prep">à</word>
  <word large_pos="Vmn—" small_pos="verb">arriver</word>
  <place>
    <word large_pos="Sp" small_pos="prep">à</word>
    <word large_pos="Da-fs-d" small_pos="det">la</word>
    <word large_pos="Ncfs" small_pos="noun">maison</word>
    <word large_pos="Sp" small_pos="prep">de</word>
  </person>
    <word large_pos="Da-fs-d" small_pos="det">la</word>
    <word large_pos="Ncfs" small_pos="noun">mère-grand</word>
  </person>
</place>
<punct ptype="semicolon">;</punct>
<person persid="4">
  <word large_pos="Pp3msn—" small_pos="pronoun">il</word>
</person>
<word large_pos="Vmip3s—" small_pos="verb">heurte</word>
<punct ptype="colon">:</punct>
<word large_pos="Npms" small_pos="noun">toc</word>
<punct ptype="comma">,</punct>
<word large_pos="Npms" small_pos="noun">toc</word>
<punct ptype="period">.</punct>
</pseudo_sentence>
</narr>
<spkr persid="3">
  <pseudo_sentence psid="20">
    <punct ptype="dash"></punct>
    <punct ptype="lquote">"</punct>
    <word large_pos="Pr-ms—" small_pos="pronoun">qui</word>
    <word large_pos="Vmip3s—" small_pos="verb">est</word>
    <place>
      <word large_pos="Rgp" small_pos="adv">là</word>
    </place>
    <punct ptype="interrogation">?</punct>
    <punct ptype="rquote">"</punct>
  </pseudo_sentence>
</spkr>
</scene>
<refrain refrid="1" occurrence="1">
  <spkr persid="4">
    <pseudo_sentence psid="21">
      <punct ptype="dash"></punct>
      <punct ptype="lquote">"</punct>
      <word large_pos="Pd-ms—" small_pos="pronoun">c'</word>
      <word large_pos="Vmip3s—" small_pos="verb">est</word>
    </person>
      <word large_pos="Ds2fsp—" small_pos="det">votre</word>
      <word large_pos="Ncfs" small_pos="noun">fille</word>

```

```

        <word large_pos="Da-ms-d" small_pos="det">le</word>
        <word large_pos="Afpms" small_pos="adj">petit</word>
        <word large_pos="Ncms" small_pos="noun">chaperon</word>
        <word large_pos="Afpms" small_pos="adj">rouge</word>
    </person>
    <punct ptype="rquote">"</punct>
</pseudo_sentence>
</spkr>
<narr>
    <pseudo_sentence psid="22">
        <punct ptype="lpar">(</punct>
        <word large_pos="Vmip3s-" small_pos="verb">dit</word>
        <person persid="4">
            <word large_pos="Da-ms-d" small_pos="det">le</word>
            <word large_pos="Ncms" small_pos="noun">loup</word>
        </person>
        <punct ptype="comma">,</punct>
        <word large_pos="Sp" small_pos="prep">en</word>
        <word large_pos="Vmn—" small_pos="verb">contrefaisant</word>
        <word large_pos="Ds3fss-" small_pos="det">sa</word>
        <word large_pos="Ncfs" small_pos="noun">voix</word>
        <punct ptype="rpar">)</punct>
    </pseudo_sentence>
</narr>
<spkr persid="4">
    <pseudo_sentence psid="23">
        <punct ptype="lquote">"</punct>
        <word large_pos="Pr-ms—" small_pos="pronoun">qui</word>
        <person persid="3">
            <word large_pos="Pp2mpa-" small_pos="pronoun">vous</word>
        </person>
        <word large_pos="Vmip3s-" small_pos="verb">apporte</word>
        <enum>
            <word large_pos="Da-fs-i" small_pos="det">une</word>
            <word large_pos="Ncfs" small_pos="noun">galette</word>
            <word large_pos="Cc" small_pos="conj">et</word>
            <word large_pos="Da-ms-i" small_pos="det">un</word>
            <word large_pos="Afpms" small_pos="adj">petit</word>
            <word large_pos="Ncms" small_pos="noun">pot</word>
            <word large_pos="Sp" small_pos="prep">de</word>
            <word large_pos="Ncms" small_pos="noun">beurre</word>
        </enum>
        <word large_pos="Cs" small_pos="conj">que</word>
        <person persid="2">
            <word large_pos="Ds1fss-" small_pos="det">ma</word>
            <word large_pos="Ncfs" small_pos="noun">mère</word>
        </person>
        <person persid="3">
            <word large_pos="Pp2mpa-" small_pos="pronoun">vous</word>
        </person>
        <word large_pos="Vmip1s-" small_pos="verb">envoie</word>
        <punct ptype="period">.</punct>
        <punct ptype="rquote">"</punct>
    </pseudo_sentence>
</spkr>
<narr>
    <pseudo_sentence psid="24">
        <person persid="3">
            <word large_pos="Da-fs-d" small_pos="det">la</word>
            <word large_pos="Afpfs" small_pos="adj">bonne</word>
            <word large_pos="Ncfs" small_pos="noun">mère-grand</word>
        </person>
        <punct ptype="comma">,</punct>
        <word large_pos="Pr-ms—" small_pos="pronoun">qui</word>
        <word large_pos="Vmii3s-" small_pos="verb">était</word>
        <place>
            <word large_pos="Sp" small_pos="prep">dans</word>
            <word large_pos="Ds3mss-" small_pos="det">son</word>
            <word large_pos="Ncms" small_pos="noun">lit</word>
        </place>
        <punct ptype="comma">,</punct>
        <word large_pos="Cc" small_pos="conj">car</word>
        <word large_pos="Pp3fsn-" small_pos="pronoun">elle</word>
        <word large_pos="Px3fs—" small_pos="pronoun">se</word>
        <word large_pos="Vmii3s-" small_pos="verb">trouvait</word>
        <word large_pos="Da-ms-i" small_pos="det">un</word>
        <word large_pos="Rgp" small_pos="adv">peu</word>
        <word large_pos="Rgp" small_pos="adv">mal</word>
    </pseudo_sentence>
</narr>

```

```

    <punct ptype="comma">,</punct>
    <person persid="4">
      <word large_pos="Pp3msd-" small_pos="pronoun">lui</word>
    </person>
    <word large_pos="Vmis3s-" small_pos="verb">cria</word>
    <punct ptype="colon">:</punct>
  </pseudo_sentence>
</narr>
<spkr persid="3">
  <pseudo_sentence psid="25">
    <punct ptype="dash">—</punct>
    <punct ptype="lquote">"</punct>
    <word large_pos="Vmip3s-" small_pos="verb">tire</word>
    <word large_pos="Da-fs-d" small_pos="det">la</word>
    <word large_pos="Ncfs" small_pos="noun">chevillette</word>
    <punct ptype="comma">,</punct>
    <word large_pos="Da-fs-d" small_pos="det">la</word>
    <word large_pos="Ncfs" small_pos="noun">bobinette</word>
    <word large_pos="Afpfs" small_pos="adj">cherra</word>
    <punct ptype="period">.</punct>
    <punct ptype="rquote">"</punct>
  </pseudo_sentence>
</spkr>
</refrain>
<scene sceneid="3">
  <narr>
    <pseudo_sentence psid="26">
      <person persid="4">
        <word large_pos="Da-ms-d" small_pos="det">le</word>
        <word large_pos="Ncms" small_pos="noun">loup</word>
      </person>
      <enum>
        <word large_pos="Vmis3s-" small_pos="verb">tira</word>
        <word large_pos="Da-fs-d" small_pos="det">la</word>
        <word large_pos="Ncfs" small_pos="noun">chevillette</word>
        <punct ptype="comma">,</punct>
        <word large_pos="Cc" small_pos="conj">et</word>
        <word large_pos="Da-fs-d" small_pos="det">la</word>
        <word large_pos="Ncfs" small_pos="noun">porte</word>
        <word large_pos="Px3fs—" small_pos="pronoun">s'</word>
        <word large_pos="Vmis3s-" small_pos="verb">ouvrit</word>
      </enum>
      <punct ptype="period">.</punct>
    </pseudo_sentence>
  </narr>
  <narr>
    <pseudo_sentence psid="27">
      <person persid="4">
        <word large_pos="Pp3msn-" small_pos="pronoun">il</word>
      </person>
      <word large_pos="Px3ms—" small_pos="pronoun">se</word>
      <word large_pos="Vmis3s-" small_pos="verb">jeta</word>
      <word large_pos="Sp" small_pos="prep">sur</word>
      <person persid="3">
        <word large_pos="Da-fs-d" small_pos="det">la</word>
        <word large_pos="Afpfs" small_pos="adj">bonne</word>
        <word large_pos="Ncfs" small_pos="noun">femme</word>
      </person>
      <punct ptype="comma">,</punct>
      <word large_pos="Cc" small_pos="conj">et</word>
      <person persid="3">
        <word large_pos="Pp3fsa-" small_pos="pronoun">la</word>
      </person>
      <word large_pos="Vmis3s-" small_pos="verb">dévora</word>
      <word large_pos="Sp" small_pos="prep">en</word>
      <word large_pos="Ncms" small_pos="noun">moins</word>
      <word large_pos="Sp" small_pos="prep">de</word>
      <word large_pos="Pi-ms—" small_pos="pronoun">rien</word>
      <punct ptype="semicolon">;</punct>
      <word large_pos="Cc" small_pos="conj">car</word>
      <word large_pos="Pp3msn-" small_pos="pronoun">il</word>
      <word large_pos="Pp3msd-" small_pos="pronoun">y</word>
      <word large_pos="Vaii3s-" small_pos="aux">avait</word>
      <word large_pos="Rgn" small_pos="adv">plus</word>
      <word large_pos="Sp" small_pos="prep">de</word>
      <word large_pos="Dk-mp—" small_pos="det">trois</word>
      <word large_pos="Ncmp" small_pos="noun">jours</word>
      <word large_pos="Cs" small_pos="conj">qu'</word>
    </pseudo_sentence>
  </narr>

```

```

<person persid="4">
  <word large_pos="Pp3msn—" small_pos="pronoun">il</word>
</person>
<word large_pos="Rpn" small_pos="adv">n'</word>
<word large_pos="Vaii3s—" small_pos="aux">avait</word>
<word large_pos="Vmps-sm" small_pos="verb">mangé</word>
<punct ptype="period">.</punct>
</pseudo_sentence>
</narr>
<narr>
  <pseudo_sentence psid="28">
    <word large_pos="Rgp" small_pos="adv">ensuite</word>
    <person persid="4">
      <word large_pos="Pp3msn—" small_pos="pronoun">il</word>
    </person>
    <word large_pos="Vmis3s—" small_pos="verb">ferma</word>
    <word large_pos="Da-fs-d" small_pos="det">la</word>
    <word large_pos="Ncfs" small_pos="noun">porte</word>
    <punct ptype="comma">,</punct>
    <word large_pos="Cc" small_pos="conj">et</word>
    <word large_pos="Px3ms—" small_pos="pronoun">s'</word>
    <word large_pos="Vmis3s—" small_pos="verb">alla</word>
    <word large_pos="Vmn—" small_pos="verb">coucher</word>
    <place>
      <word large_pos="Sp" small_pos="prep">dans</word>
      <word large_pos="Da-ms-d" small_pos="det">le</word>
      <word large_pos="Ncms" small_pos="noun">lit</word>
      <word large_pos="Sp" small_pos="prep">de</word>
    </place>
    <person persid="3">
      <word large_pos="Da-fs-d" small_pos="det">la</word>
      <word large_pos="Ncfs" small_pos="noun">mère-grand</word>
    </person>
    </place>
    <punct ptype="comma">,</punct>
    <word large_pos="Sp" small_pos="prep">en</word>
    <word large_pos="Vmpp—" small_pos="verb">attendait</word>
    <person persid="1">
      <word large_pos="Da-ms-d" small_pos="det">le</word>
      <word large_pos="Afms" small_pos="adj">petit</word>
      <word large_pos="Ncms" small_pos="noun">chaperon</word>
      <word large_pos="Afms" small_pos="adj">rouge</word>
    </person>
    <punct ptype="comma">,</punct>
    <word large_pos="Pr-ms—" small_pos="pronoun">qui</word>
    <time>
      <word large_pos="Di-ms—" small_pos="det">quelque</word>
      <word large_pos="Ncms" small_pos="noun">temps</word>
      <word large_pos="Sp" small_pos="prep">après</word>
    </time>
    <word large_pos="Vmis3s—" small_pos="verb">vint</word>
    <word large_pos="Vmn—" small_pos="verb">heurter</word>
    <word large_pos="Sp" small_pos="prep">à</word>
    <word large_pos="Da-fs-d" small_pos="det">la</word>
    <word large_pos="Ncfs" small_pos="noun">porte</word>
    <punct ptype="period">.</punct>
  </pseudo_sentence>
</narr>
<narr>
  <pseudo_sentence psid="29">
    <word large_pos="Npms" small_pos="noun">toc</word>
    <punct ptype="comma">,</punct>
    <word large_pos="Npms" small_pos="noun">toc</word>
    <punct ptype="period">.</punct>
  </pseudo_sentence>
</narr>
<spkr persid="4">
  <pseudo_sentence psid="30">
    <punct ptype="dash"></punct>
    <punct ptype="lquote">"</punct>
    <word large_pos="Pr-ms—" small_pos="pronoun">qui</word>
    <word large_pos="Vmip3s—" small_pos="verb">est</word>
    <place>
      <word large_pos="Rgp" small_pos="adv">là</word>
    </place>
    <punct ptype="interrogation">?</punct>
    <punct ptype="rquote">"</punct>
    <punct ptype="period">.</punct>
  </pseudo_sentence>

```

```

</spkr>
<narr>
  <pseudo.sentence psid="31">
    <person persid="1">
      <word large_pos="Da-ms-d" small_pos="det">le</word>
      <word large_pos="Afpms" small_pos="adj">petit</word>
      <word large_pos="Ncms" small_pos="noun">chaperon</word>
      <word large_pos="Afpms" small_pos="adj">rouge</word>
    </person>
    <punct ptype="comma">,</punct>
    <word large_pos="Pr-ms—" small_pos="pronoun">qui</word>
    <word large_pos="Vmis3s—" small_pos="verb">entendit</word>
    <word large_pos="Da-fs-d" small_pos="det">la</word>
    <word large_pos="Afpfs" small_pos="adj">grosse</word>
    <word large_pos="Ncfs" small_pos="noun">voix</word>
    <person persid="4">
      <word large_pos="Sp+Da-ms-d" small_pos="prep">du</word>
      <word large_pos="Ncms" small_pos="noun">loup</word>
    </person>
    <punct ptype="comma">,</punct>
    <word large_pos="Vmis3s—" small_pos="verb">eut</word>
    <word large_pos="Ncfs" small_pos="noun">peur</word>
    <word large_pos="Afpfs" small_pos="adj">d'abord</word>
    <punct ptype="comma">,</punct>
    <word large_pos="Cc" small_pos="conj">mais</word>
    <word large_pos="Vmpp—" small_pos="verb">croyant</word>
    <word large_pos="Cs" small_pos="conj">que</word>
    <person persid="3">
      <word large_pos="Ds3fss—" small_pos="det">sa</word>
      <word large_pos="Ncfs" small_pos="noun">mère-grand</word>
    </person>
    <word large_pos="Vaii3s—" small_pos="aux">était</word>
    <word large_pos="Vmpps-sm" small_pos="verb">enrhumée</word>
    <punct ptype="comma">,</punct>
    <word large_pos="Vmis3s—" small_pos="verb">répondit</word>
    <punct ptype="colon">:</punct>
  </pseudo.sentence>
</narr>
</scene>
<refrain refrid="1" occurrence="2">
  <spkr persid="1">
    <pseudo.sentence psid="32">
      <punct ptype="dash"></punct>
      <punct ptype="lquote">"</punct>
      <word large_pos="Pd-ms—" small_pos="pronoun">c'</word>
      <word large_pos="Vmip3s—" small_pos="verb">est</word>
      <person persid="1">
        <word large_pos="Ds2fsp—" small_pos="det">votre</word>
        <word large_pos="Ncfs" small_pos="noun">fille</word>
        <word large_pos="Da-ms-d" small_pos="det">le</word>
        <word large_pos="Afpms" small_pos="adj">petit</word>
        <word large_pos="Ncms" small_pos="noun">chaperon</word>
        <word large_pos="Afpms" small_pos="adj">rouge</word>
      </person>
      <punct ptype="comma">,</punct>
      <word large_pos="Pr-ms—" small_pos="pronoun">qui</word>
      <word large_pos="Pp2mpa—" small_pos="pronoun">vous</word>
      <word large_pos="Vmip3s—" small_pos="verb">apporte</word>
      <enum>
        <word large_pos="Da-fs-i" small_pos="det">une</word>
        <word large_pos="Ncfs" small_pos="noun">galette</word>
        <word large_pos="Cc" small_pos="conj">et</word>
        <word large_pos="Da-ms-i" small_pos="det">un</word>
        <word large_pos="Afpms" small_pos="adj">petit</word>
        <word large_pos="Ncms" small_pos="noun">pot</word>
        <word large_pos="Sp" small_pos="prep">de</word>
        <word large_pos="Ncms" small_pos="noun">beurre</word>
      </enum>
      <word large_pos="Cs" small_pos="conj">que</word>
    </pseudo.sentence>
    <person persid="2">
      <word large_pos="Ds1fss—" small_pos="det">ma</word>
      <word large_pos="Ncfs" small_pos="noun">mère</word>
    </person>
    <person persid="3">
      <word large_pos="Pp2mpa—" small_pos="pronoun">vous</word>
    </person>
    <word large_pos="Vmip1s—" small_pos="verb">envoie</word>
    <punct ptype="period">.</punct>
  </spkr>
</refrain>

```

```

    <punct ptype="rquote"></punct>
  </pseudo_sentence>
</spkr>
<narr>
  <pseudo_sentence psid="33">
    <person persid="4">
      <word large_pos="Da-ms-d" small_pos="det">le</word>
      <word large_pos="Ncms" small_pos="noun">loup</word>
    </person>
    <person persid="1">
      <word large_pos="Pp3msd-" small_pos="pronoun">lui</word>
    </person>
    <word large_pos="Vmis3s-" small_pos="verb">cria</word>
    <punct ptype="comma">,</punct>
    <word large_pos="Sp" small_pos="prep">en</word>
    <word large_pos="Vmn—" small_pos="verb">adoucissant</word>
    <word large_pos="Da-ms-i" small_pos="det">un</word>
    <word large_pos="Rgp" small_pos="adv">peu</word>
    <word large_pos="Ds3fss-" small_pos="det">sa</word>
    <word large_pos="Ncfs" small_pos="noun">voix</word>
    <punct ptype="colon">:</punct>
  </pseudo_sentence>
</narr>
<spkr persid="4">
  <pseudo_sentence psid="34">
    <punct ptype="dash"></punct>
    <punct ptype="lquote"></punct>
    <word large_pos="Vmip3s-" small_pos="verb">tire</word>
    <word large_pos="Da-fs-d" small_pos="det">la</word>
    <word large_pos="Ncfs" small_pos="noun">chevillette</word>
    <punct ptype="comma">,</punct>
    <word large_pos="Da-fs-d" small_pos="det">la</word>
    <word large_pos="Ncfs" small_pos="noun">bobinette</word>
    <word large_pos="Afpfs" small_pos="adj">cherra</word>
    <punct ptype="period">.</punct>
    <punct ptype="rquote"></punct>
  </pseudo_sentence>
</spkr>
</refrain>
<scene sceneid="4">
  <narr>
    <pseudo_sentence psid="35">
      <person persid="1">
        <word large_pos="Da-ms-d" small_pos="det">le</word>
        <word large_pos="Afpms" small_pos="adj">petit</word>
        <word large_pos="Ncms" small_pos="noun">chaperon</word>
        <word large_pos="Afpms" small_pos="adj">rouge</word>
      </person>
      <word large_pos="Vmis3s-" small_pos="verb">tira</word>
      <word large_pos="Da-fs-d" small_pos="det">la</word>
      <word large_pos="Ncfs" small_pos="noun">chevillette</word>
      <punct ptype="comma">,</punct>
      <word large_pos="Cc" small_pos="conj">et</word>
      <word large_pos="Da-fs-d" small_pos="det">la</word>
      <word large_pos="Ncfs" small_pos="noun">porte</word>
      <word large_pos="Px3fs—" small_pos="pronoun">s'</word>
      <word large_pos="Vmis3s-" small_pos="verb">ouvrit</word>
      <punct ptype="period">.</punct>
    </pseudo_sentence>
    <pseudo_sentence psid="36">
      <person persid="4">
        <word large_pos="Da-ms-d" small_pos="det">le</word>
        <word large_pos="Ncms" small_pos="noun">loup</word>
      </person>
      <punct ptype="comma">,</punct>
      <person persid="1">
        <word large_pos="Pp3fsa-" small_pos="pronoun">la</word>
      </person>
      <word large_pos="Vmpp—" small_pos="verb">voyant</word>
      <word large_pos="Vmn—" small_pos="verb">entrer</word>
      <punct ptype="comma">,</punct>
      <person persid="1">
        <word large_pos="Pp3msd-" small_pos="pronoun">lui</word>
      </person>
      <word large_pos="Vmip3s-" small_pos="verb">dit</word>
      <word large_pos="Sp" small_pos="prep">en</word>
      <word large_pos="Px3ms—" small_pos="pronoun">se</word>
      <word large_pos="Vmpp—" small_pos="verb">cachant</word>

```

```

    <place>
    <word large_pos="Sp" small_pos="prep">dans</word>
    <word large_pos="Da-ms-d" small_pos="det">le</word>
    <word large_pos="Ncms" small_pos="noun">lit</word>
    <word large_pos="Sp" small_pos="prep">sous</word>
    <word large_pos="Da-fs-d" small_pos="det">la</word>
    <word large_pos="Ncfs" small_pos="noun">couverture</word>
  </place>
  <punct ptype="colon">:</punct>
</pseudo_sentence>
</narr>
<spkr persid="4">
  <pseudo_sentence psid="37">
    <punct ptype="dash"></punct>
    <punct ptype="lquote">"</punct>
    <word large_pos="Vmmp2s-" small_pos="verb">mets</word>
    <enum>
      <word large_pos="Da-fs-d" small_pos="det">la</word>
      <word large_pos="Ncfs" small_pos="noun">galette</word>
      <word large_pos="Cc" small_pos="conj">et</word>
      <word large_pos="Da-ms-d" small_pos="det">le</word>
      <word large_pos="Afpms" small_pos="adj">petit</word>
      <word large_pos="Ncms" small_pos="noun">pot</word>
      <word large_pos="Sp" small_pos="prep">de</word>
      <word large_pos="Ncms" small_pos="noun">beurre</word>
    </enum>
    <word large_pos="Sp" small_pos="prep">sur</word>
    <word large_pos="Da-fs-d" small_pos="det">la</word>
    <word large_pos="Ncfs" small_pos="noun">huche</word>
    <punct ptype="comma">,</punct>
    <word large_pos="Cc" small_pos="conj">et</word>
    <word large_pos="Vmmp2s-" small_pos="verb">viens</word>
    <person persid="1">
      <word large_pos="Pp2msd-" small_pos="pronoun">te</word>
    </person>
    <word large_pos="Vmn—" small_pos="verb">coucher</word>
    <word large_pos="Sp" small_pos="prep">avec</word>
    <person persid="4">
      <word large_pos="Pp1ms—" small_pos="pronoun">moi</word>
    </person>
    <punct ptype="period">.</punct>
    <punct ptype="rquote">"</punct>
  </pseudo_sentence>
</spkr>
</narr>
<pseudo_sentence psid="38">
  <person persid="1">
    <word large_pos="Da-ms-d" small_pos="det">le</word>
    <word large_pos="Afpms" small_pos="adj">petit</word>
    <word large_pos="Ncms" small_pos="noun">chaperon</word>
    <word large_pos="Afpms" small_pos="adj">rouge</word>
  </person>
  <word large_pos="Px3ms—" small_pos="pronoun">se</word>
  <word large_pos="Vmn—" small_pos="verb">déshabille</word>
  <punct ptype="comma">,</punct>
  <word large_pos="Cc" small_pos="conj">et</word>
  <word large_pos="Vmip3s-" small_pos="verb">va</word>
  <person persid="1">
    <word large_pos="Px3ms—" small_pos="pronoun">se</word>
  </person>
  <word large_pos="Vmn—" small_pos="verb">mettre</word>
  <place>
    <word large_pos="Sp" small_pos="prep">dans</word>
    <word large_pos="Da-ms-d" small_pos="det">le</word>
    <word large_pos="Ncms" small_pos="noun">lit</word>
  </place>
  <punct ptype="comma">,</punct>
  <word large_pos="Pr-ms—" small_pos="pronoun">où</word>
  <person persid="1">
    <word large_pos="Pp3fsn-" small_pos="pronoun">elle</word>
  </person>
  <word large_pos="Vais3s-" small_pos="aux">fut</word>
  <word large_pos="Rgp" small_pos="adv">bien</word>
  <word large_pos="Vmips-sf" small_pos="verb">étonnée</word>
  <word large_pos="Sp" small_pos="prep">de</word>
  <word large_pos="Vmn—" small_pos="verb">voir</word>
  <word large_pos="Rxp" small_pos="adv">comment</word>
  <person persid="3">

```



```

    <word large_pos="Ds3fss-" small_pos="det">sa</word>
    <word large_pos="Ncfs" small_pos="noun">mère-grand</word>
  </person>
  <word large_pos="Vaii3s-" small_pos="aux">était</word>
  <word large_pos="Vm3s-sf" small_pos="verb">faite</word>
  <word large_pos="Sp" small_pos="prep">en</word>
  <word large_pos="Ds3mss-" small_pos="det">son</word>
  <word large_pos="Afpms" small_pos="adj">déshabillé</word>
  <punct ptype="period">.</punct>
</pseudo_sentence>
</narr>
<narr>
  <pseudo_sentence psid="39">
    <person persid="1">
      <word large_pos="Pp3fsn-" small_pos="pronoun">elle</word>
    </person>
    <person persid="3">
      <word large_pos="Pp3msd-" small_pos="pronoun">lui</word>
    </person>
    <word large_pos="Vmip3s-" small_pos="verb">dit</word>
    <punct ptype="colon">:</punct>
  </pseudo_sentence>
</narr>
<spkr persid="1">
  <pseudo_sentence psid="40">
    <punct ptype="dash"></punct>
    <punct ptype="lquote">"</punct>
    <person persid="3">
      <word large_pos="Dslfss-" small_pos="det">ma</word>
      <word large_pos="Ncfs" small_pos="noun">mère-grand</word>
    </person>
    <word large_pos="Cs" small_pos="conj">que</word>
    <person persid="3">
      <word large_pos="Pp2mpn-" small_pos="pronoun">vous</word>
    </person>
    <word large_pos="Vmip2p-" small_pos="verb">avez</word>
    <word large_pos="Sp" small_pos="prep">de</word>
    <word large_pos="Afpmp" small_pos="adj">grands</word>
    <word large_pos="Ncmp" small_pos="noun">bras</word>
    <punct ptype="exclamation">!"</punct>
    <punct ptype="rquote">"</punct>
  </pseudo_sentence>
</spkr>
<spkr persid="4">
  <pseudo_sentence psid="41">
    <punct ptype="dash"></punct>
    <punct ptype="lquote">"</punct>
    <word large_pos="Pd-ms—" small_pos="pronoun">c'</word>
    <word large_pos="Vmip3s-" small_pos="verb">est</word>
    <word large_pos="Sp" small_pos="prep">pour</word>
    <word large_pos="Rgc" small_pos="adv">mieux</word>
    <person persid="1">
      <word large_pos="Pp2msa-" small_pos="pronoun">t'</word>
    </person>
    <word large_pos="Vmn—" small_pos="verb">embrasser</word>
    <person persid="1">
      <word large_pos="Dslfss-" small_pos="det">ma</word>
      <word large_pos="Ncfs" small_pos="noun">fille</word>
    </person>
    <punct ptype="period">.</punct>
    <punct ptype="rquote">"</punct>
  </pseudo_sentence>
</spkr>
<spkr persid="1">
  <pseudo_sentence psid="42">
    <punct ptype="dash"></punct>
    <punct ptype="lquote">"</punct>
    <person persid="3">
      <word large_pos="Dslfss-" small_pos="det">ma</word>
      <word large_pos="Ncfs" small_pos="noun">mère-grand</word>
    </person>
    <word large_pos="Cs" small_pos="conj">que</word>
    <person persid="3">
      <word large_pos="Pp2mpn-" small_pos="pronoun">vous</word>
    </person>
    <word large_pos="Vmip2p-" small_pos="verb">avez</word>
    <word large_pos="Sp" small_pos="prep">de</word>
    <word large_pos="Afpfp" small_pos="adj">grandes</word>
  </pseudo_sentence>
</spkr>

```

```

    <word large_pos="Ncfp" small_pos="noun">jambes</word>
    <punct ptype="exclamation">!</punct>
    <punct ptype="rquote">"</punct>
  </pseudo_sentence>
</spkr>
<spkr persid="4">
  <pseudo_sentence psid="43">
    <punct ptype="dash"></punct>
    <punct ptype="lquote">"</punct>
    <word large_pos="Pd-ms—" small_pos="pronoun">c'</word>
    <word large_pos="Vmip3s—" small_pos="verb">est</word>
    <word large_pos="Sp" small_pos="prep">pour</word>
    <word large_pos="Rgc" small_pos="adv">mieux</word>
    <word large_pos="Vmn—" small_pos="verb">courir</word>
    <person persid="1">
      <word large_pos="Dslmss—" small_pos="det">mon</word>
      <word large_pos="Ncms" small_pos="noun">enfant</word>
    </person>
    <punct ptype="period">.</punct>
    <punct ptype="rquote">"</punct>
  </pseudo_sentence>
</spkr>
<spkr persid="1">
  <pseudo_sentence psid="44">
    <punct ptype="dash"></punct>
    <punct ptype="lquote">"</punct>
    <person persid="3">
      <word large_pos="Dslfss—" small_pos="det">ma</word>
      <word large_pos="Ncfs" small_pos="noun">mère-grand</word>
    </person>
    <word large_pos="Cs" small_pos="conj">que</word>
    <person persid="3">
      <word large_pos="Pp2mpn—" small_pos="pronoun">vous</word>
    </person>
    <word large_pos="Vmip2p—" small_pos="verb">avez</word>
    <word large_pos="Sp" small_pos="prep">de</word>
    <word large_pos="Afpfp" small_pos="adj">grandes</word>
    <word large_pos="Ncfp" small_pos="noun">oreilles</word>
    <punct ptype="exclamation">!</punct>
    <punct ptype="rquote">"</punct>
  </pseudo_sentence>
</spkr>
<spkr persid="4">
  <pseudo_sentence psid="45">
    <punct ptype="dash"></punct>
    <punct ptype="lquote">"</punct>
    <word large_pos="Pd-ms—" small_pos="pronoun">c'</word>
    <word large_pos="Vmip3s—" small_pos="verb">est</word>
    <word large_pos="Sp" small_pos="prep">pour</word>
    <word large_pos="Rgc" small_pos="adv">mieux</word>
    <word large_pos="Vmn—" small_pos="verb">écouter</word>
    <person persid="1">
      <word large_pos="Dslmss—" small_pos="det">mon</word>
      <word large_pos="Ncms" small_pos="noun">enfant</word>
    </person>
    <punct ptype="period">.</punct>
    <punct ptype="rquote">"</punct>
  </pseudo_sentence>
</spkr>
<spkr persid="1">
  <pseudo_sentence psid="46">
    <punct ptype="dash"></punct>
    <punct ptype="lquote">"</punct>
    <person persid="3">
      <word large_pos="Dslfss—" small_pos="det">ma</word>
      <word large_pos="Ncfs" small_pos="noun">mère-grand</word>
    </person>
    <word large_pos="Cs" small_pos="conj">que</word>
    <person persid="3">
      <word large_pos="Pp2mpn—" small_pos="pronoun">vous</word>
    </person>
    <word large_pos="Vmip2p—" small_pos="verb">avez</word>
    <word large_pos="Sp" small_pos="prep">de</word>
    <word large_pos="Afpmp" small_pos="adj">grands</word>
    <word large_pos="Ncmp" small_pos="noun">yeux</word>
    <punct ptype="exclamation">!</punct>
    <punct ptype="rquote">"</punct>
  </pseudo_sentence>

```

```

</spkr>
<spkr persid="4">
  <pseudo.sentence psid="47">
    <punct ptype="dash"></punct>
    <punct ptype="lquote"></punct>
    <word large_pos="Pd-ms—" small_pos="pronoun">c'</word>
    <word large_pos="Vmip3s—" small_pos="verb">est</word>
    <word large_pos="Sp" small_pos="prep">pour</word>
    <word large_pos="Rgc" small_pos="adv">mieux</word>
    <word large_pos="Vmn—" small_pos="verb">voir</word>
    <person persid="1">
      <word large_pos="Dslmss—" small_pos="det">mon</word>
      <word large_pos="Ncms" small_pos="noun">enfant</word>
    </person>
    <punct ptype="period">.</punct>
    <punct ptype="rquote"></punct>
  </pseudo.sentence>
</spkr>
<spkr persid="1">
  <pseudo.sentence psid="48">
    <punct ptype="dash"></punct>
    <punct ptype="lquote"></punct>
    <person persid="3">
      <word large_pos="Dslfss—" small_pos="det">ma</word>
      <word large_pos="Ncfs" small_pos="noun">mère-grand</word>
    </person>
    <word large_pos="Cs" small_pos="conj">que</word>
    <person persid="1">
      <word large_pos="Pp2mpn—" small_pos="pronoun">vous</word>
    </person>
    <word large_pos="Vmip2p—" small_pos="verb">avez</word>
    <word large_pos="Sp" small_pos="prep">de</word>
    <word large_pos="Afpfp" small_pos="adj">grandes</word>
    <word large_pos="Ncfp" small_pos="noun">dents</word>
    <punct ptype="exclamation">!</punct>
    <punct ptype="rquote"></punct>
  </pseudo.sentence>
</spkr>
<spkr persid="4">
  <pseudo.sentence psid="49">
    <punct ptype="dash"></punct>
    <punct ptype="lquote"></punct>
    <word large_pos="Pd-ms—" small_pos="pronoun">c'</word>
    <word large_pos="Vmip3s—" small_pos="verb">est</word>
    <word large_pos="Sp" small_pos="prep">pour</word>
    <person persid="1">
      <word large_pos="Pp2msd—" small_pos="pronoun">te</word>
    </person>
    <word large_pos="Vmn—" small_pos="verb">manger</word>
    <punct ptype="period">.</punct>
    <punct ptype="rquote"></punct>
  </pseudo.sentence>
</spkr>
</scene>
<epilogue>
  <narr>
    <pseudo.sentence psid="50">
      <word large_pos="Cc" small_pos="conj">et</word>
      <word large_pos="Sp" small_pos="prep">en</word>
      <word large_pos="Vmpp—" small_pos="verb">disant</word>
      <word large_pos="Dd-mp—" small_pos="det">ces</word>
      <word large_pos="Ncmp" small_pos="noun">mots</word>
      <punct ptype="comma">,</punct>
      <person persid="4">
        <word large_pos="Da-ms-d" small_pos="det">le</word>
        <word large_pos="Afpms" small_pos="adj">méchant</word>
        <word large_pos="Ncms" small_pos="noun">loup</word>
      </person>
      <enum>
        <word large_pos="Px3ms—" small_pos="pronoun">se</word>
        <word large_pos="Vmis3s—" small_pos="verb">jeta</word>
        <word large_pos="Sp" small_pos="prep">sur</word>
        <person persid="1">
          <word large_pos="Da-ms-d" small_pos="det">le</word>
          <word large_pos="Afpms" small_pos="adj">petit</word>
          <word large_pos="Ncms" small_pos="noun">chaperon</word>
          <word large_pos="Afpms" small_pos="adj">rouge</word>
        </person>
      </enum>
    </pseudo.sentence>
  </narr>
</epilogue>

```

```
<punct ptype="comma">,</punct>
<word large_pos="Cc" small_pos="conj">et</word>
<person persid="1">
  <word large_pos="Pp3fsa-" small_pos="pronoun">la</word>
</person>
<word large_pos="Vmis3s-" small_pos="verb">mangea</word>
</enum>
<punct ptype="period">.</punct>
</pseudo_sentence>
</narr>
</epilogue>
</gvlex_tale>
```



## Annexe C

# Modélisation CRF des tours de parole

Cette annexe fournit le détail des caractéristiques des données ayant permis d’obtenir les meilleurs résultats sur la tâche de segmentation en tours de parole basée sur des CRF détaillée dans la section 3.3.

### C.1 Données d’entrée

Le meilleur ensemble de descripteurs utilisé pour la tâche de segmentation en tours de parole consiste à considérer des observations à 13 dimensions. La première dimension correspond au texte brut, les neuf suivantes correspondent aux parties du discours obtenues à l’aide du système (Allauzen and Bonneau-Maynard, 2008) séparées en descripteurs atomiques, les suivantes correspondent à la présence de majuscule, à la partie du discours obtenue à l’aide de TreeTagger (Stein and Schmid, 1995), et la dernière correspond à la distinction mot/signe de ponctuation. Le dernier champ correspond à l’annotation attendue. L’utilisation de ces descripteurs est illustré sur le conte “Le petit chaperon rouge”, détaillé ci-dessous.

```
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos BEGIN_narr
petit A Af Ams Ap petit petit petit petit Afpms Min ADJ pos INSIDE_narr
chaperon N Nc Nms chaperon chaperon chaperon chaperon chaperon Ncms Min NOM pos INSIDE_narr
rouge A Af Ams Ap rouge rouge rouge rouge Afpms Min ADJ pos INSIDE_narr
\n\n - - - - - Min - punctuation INSIDE_narr
il P Pp P3ms il P- il Pn il Pp3msn- Min PRO:PER pos INSIDE_narr
était V Vm V3-s était était Vii était était Vmii3s- Min VER:impf pos INSIDE_narr
une D Da D-fs une D- une une Di Da-fs-i Min DET:ART pos INSIDE_narr
fois N Nc Nfs fois fois fois fois fois Ncfs Min NOM pos INSIDE_narr
une D Da D-fs une D- une une Di Da-fs-i Min DET:ART pos INSIDE_narr
petite A Af Afs Ap petite petite petite petite Afpfs Min ADJ pos INSIDE_narr
fille N Nc Nfs fille fille fille fille fille Ncfs Min NOM pos INSIDE_narr
de S Sp de de de de de de de Sp Min PRP pos INSIDE_narr
village N Nc Nms village village village village village Ncms Min NOM pos INSIDE_narr
, F , , , , , F Min PUN punctuation INSIDE_narr
la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos INSIDE_narr
plus R Rg Rc plus plus plus plus plus Rgc Min ADV pos INSIDE_narr
jolie A Af Afs Ap jolie jolie jolie jolie Afpfs Min ADJ pos INSIDE_narr
qu' C Cs qu' qu' qu' qu' qu' qu' qu' Cs Min PRO:REL pos INSIDE_narr
on P Pp P3ms on P- on Pn on Pp3msn- Min PRO:PER pos INSIDE_narr
eût V Va V3-s eût eût Vsi eût eût Vasi3s- Min VER:subi pos INSIDE_narr
su V Vm V-ms su su Vps su su Vmps-sm Min VER:pper pos INSIDE_narr
voir V Vm V- voir voir Vn- voir voir Vmn- Min VER:infi pos INSIDE_narr
```

```

; F ; ; ; ; ; ; F Min PUN punctuation INSIDE_narr
sa D Ds D3fs sa Ds sa sa D- Ds3fss- Min DET:POS pos INSIDE_narr
mère N Nc Nfs mère mère mère mère Ncfs Min NOM pos INSIDE_narr
en P Pp P3ms en P- en Pd en Pp3msd- Min PRO:PER pos INSIDE_narr
était V Vm V3-s était était Vii était était Vmii3s- Min VER:impf pos INSIDE_narr
folle A Af Afs Ap folle folle folle folle Afpfs Min ADJ pos INSIDE_narr
, F , , , , , , F Min PUN punctuation INSIDE_narr
et C Cc et et et et et et Cc Min KON pos INSIDE_narr
sa D Ds D3fs sa Ds sa sa D- Ds3fss- Min DET:POS pos INSIDE_narr
mère-grand N Nc Nfs mère-grand mère-grand mère-grand mère-grand Ncfs Min NOM pos INSIDE_narr
plus R Rg Rc plus plus plus plus plus plus Rgc Min ADV pos INSIDE_narr
folle A Af Afs Ap folle folle folle folle Afpfs Min ADJ pos INSIDE_narr
encore R Rg Rp encore encore encore encore Rgp Min ADV pos INSIDE_narr
. F . . . . . F Min SENT punctuation INSIDE_narr
\n - - - - - Min - punctuation INSIDE_narr
cette D Dd D-fs cette D- cette cette D- Dd-fs- Min PRO:DEM pos INSIDE_narr
bonne A Af Afs Ap bonne bonne bonne bonne Afpfs Min ADJ pos INSIDE_narr
femme N Nc Nfs femme femme femme femme femme Ncfs Min NOM pos INSIDE_narr
lui P Pp P3ms lui P- lui Pd lui Pp3msd- Min PRO:PER pos INSIDE_narr
fit V Vm V3-s fit fit Vis fit fit Vmis3s- Min VER:simp pos INSIDE_narr
faire V Vm V- faire faire Vn- faire faire Vmn- Min VER:infi pos INSIDE_narr
un D Da D-ms un D- un un Di Da-ms-i Min DET:ART pos INSIDE_narr
petit A Af Ams Ap petit petit petit petit Afpms Min ADJ pos INSIDE_narr
chaperon N Nc Nms chaperon chaperon chaperon chaperon Ncms Min NOM pos INSIDE_narr
rouge A Af Ams Ap rouge rouge rouge rouge Afpms Min ADJ pos INSIDE_narr
, F , , , , , , F Min PUN punctuation INSIDE_narr
qui P Pr P-ms qui P- qui P- qui Pr-ms- Min PRO:REL pos INSIDE_narr
lui P Pp P3ms lui P- lui Pd lui Pp3msd- Min PRO:PER pos INSIDE_narr
seyait V Vm V3-s seyait seyait Vip seyait seyait Vmip3s- Min VER:impf pos INSIDE_narr
si R Rg Rp si si si si si Rgp Min ADV pos INSIDE_narr
bien R Rg Rp bien bien bien bien Rgp Min ADV pos INSIDE_narr
, F , , , , , , F Min PUN punctuation INSIDE_narr
que C Cs que que que que que que Cs Min PRO:REL pos INSIDE_narr
partout R Rg Rp partout partout partout partout Rgp Min ADV pos INSIDE_narr
on P Pp P3ms on P- on Pn on Pp3msn- Min PRO:PER pos INSIDE_narr
l' P Pp P3ms l' P- l' Pa l' Pp3msa- Min PRO:PER pos INSIDE_narr
appelait V Vm V3-s appelait appelait Vii appelait appelait Vmii3s- Min VER:impf pos INSIDE_narr
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos INSIDE_narr
petit A Af Ams Ap petit petit petit petit Afpms Min ADJ pos INSIDE_narr
chaperon N Nc Nms chaperon chaperon chaperon chaperon Ncms Min NOM pos INSIDE_narr
rouge A Af Ams Ap rouge rouge rouge rouge Afpms Min ADJ pos INSIDE_narr
. F . . . . . F Min SENT punctuation INSIDE_narr
\n\n - - - - - Min - punctuation INSIDE_narr
un D Da D-ms un D- un un Di Da-ms-i Min DET:ART pos INSIDE_narr
jour N Nc Nms jour jour jour jour jour Ncms Min NOM pos INSIDE_narr
sa D Ds D3fs sa Ds sa sa D- Ds3fss- Min DET:POS pos INSIDE_narr
mère N Nc Nfs mère mère mère mère mère Ncfs Min NOM pos INSIDE_narr
, F , , , , , , F Min PUN punctuation INSIDE_narr
ayant V Va V- ayant ayant Vpp ayant ayant Vapp- Min VER:ppe pos INSIDE_narr
cuit V Vm V-ms cuit cuit Vps cuit cuit Vmps-sm Min ADJ pos INSIDE_narr
et C Cc et et et et et et Cc Min KON pos INSIDE_narr
fait V Vm V-ms fait fait fait Vps fait fait Vmps-sm Min VER:ppe pos INSIDE_narr
des S+D Sp+Da des+D-mp des+des des+des des+des des+Dd Sp+Da-mp-d Min PRP:det pos INSIDE_narr
galettes N Nc Nmp galettes galettes galettes galettes galettes Ncmp Min NOM pos INSIDE_narr
, F , , , , , , F Min PUN punctuation INSIDE_narr
lui P Pp P3ms lui P- lui Pd lui Pp3msd- Min PRO:PER pos INSIDE_narr
dit V Vm V3-s dit dit Vip dit dit Vmip3s- Min VER:pres pos INSIDE_narr
: F : : : : : : F Min PUN punctuation INSIDE_narr
\n - - - - - Min - punctuation INSIDE_narr
- F - - - - - F Min PUN punctuation BEGIN_spkr
" F " " " " " " F Min PUN:cit punctuation INSIDE_spkr
va V Vm V3-s va va Vip va va Vmip3s- Min VER:pres pos INSIDE_spkr
voir V Vm V- voir voir Vn- voir voir Vmn- Min VER:infi pos INSIDE_spkr
comment R Rx Rp comment comment comment comment Rxp Min ADV pos INSIDE_spkr
se P Px P3ms se P- se P- se Px3ms- Min PRO:PER pos INSIDE_spkr
porte V Vm V3-s porte porte Vip porte porte Vmip3s- Min VER:pres pos INSIDE_spkr
ta D Ds D2fs ta Ds ta ta D- Ds2fss- Min DET:POS pos INSIDE_spkr
mère-grand N Nc Nfs mère-grand mère-grand mère-grand mère-grand Ncfs Min NOM pos INSIDE_spkr
, F , , , , , , F Min PUN punctuation INSIDE_spkr
car C Cc car car car car Cc Min KON pos INSIDE_spkr
on P Pp P3ms on P- on Pn on Pp3msn- Min PRO:PER pos INSIDE_spkr
m' P Pp P1ms m' P- m' Pa m' Pp1msa- Min PRO:PER pos INSIDE_spkr
a V Va V3-s a a Vip a a Vaip3s- Min VER:pres pos INSIDE_spkr
dit V Vm V-ms dit dit Vps dit dit Vmps-sm Min VER:ppe pos INSIDE_spkr
qu' C Cs qu' qu' qu' qu' qu' qu' Cs Min KON pos INSIDE_spkr
elle P Pp P3fs elle P- elle Pn elle Pp3fsn- Min PRO:PER pos INSIDE_spkr
était V Vm V3-s était était Vii était était Vmii3s- Min VER:impf pos INSIDE_spkr
malade A Af Ams Ap malade malade malade malade Afpms Min ADJ pos INSIDE_spkr

```

## C.1. Données d'entrée

, F , , , , , , F Min PUN punctuation INSIDE\_spkr  
 porte V Vm V3-s porte porte Vip porte porte Vmip3s- Min VER:pres pos INSIDE\_spkr  
 -lui P Pp P3ms -lui P- -lui Pd -lui Pp3msd- Min PRO:PER pos INSIDE\_spkr  
 une D Da D-fs une D- une une Di Da-fs-i Min DET:ART pos INSIDE\_spkr  
 galette N Nc Nfs galette galette galette galette galette Ncfs Min NOM pos INSIDE\_spkr  
 et C Cc et et et et et Cc Min KON pos INSIDE\_spkr  
 ce D Dd D-ms ce D- ce ce D- Dd-ms- Min PRO:DEM pos INSIDE\_spkr  
 petit A Af Ams Ap petit petit petit petit Afpms Min ADJ pos INSIDE\_spkr  
 pot N Nc Nms pot pot pot pot pot Ncms Min NOM pos INSIDE\_spkr  
 de S Sp de de de de de Sp Min PRP pos INSIDE\_spkr  
 beurre N Nc Nms beurre beurre beurre beurre beurre Ncms Min NOM pos INSIDE\_spkr  
 . F . . . . . F Min SENT punctuation INSIDE\_spkr  
 " F " " " " " " F Min PUN:cit punctuation INSIDE\_spkr  
 \n - - - - - Min - punctuation INSIDE\_spkr  
 le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos BEGIN\_narr  
 petit A Af Ams Ap petit petit petit petit Afpms Min ADJ pos INSIDE\_narr  
 chaperon N Nc Nms chaperon chaperon chaperon chaperon chaperon Ncms Min NOM pos INSIDE\_narr  
 rouge A Af Ams Ap rouge rouge rouge rouge Afpms Min ADJ pos INSIDE\_narr  
 parti V Vm V3-s parti parti Vis parti parti Vmis3s- Min VER:simp pos INSIDE\_narr  
 aussitôt R Rg Rp aussitôt aussitôt aussitôt aussitôt aussitôt Rgp Min ADV pos INSIDE\_narr  
 pour S Sp pour pour pour pour pour Sp Min PRP pos INSIDE\_narr  
 aller V Vm V- aller aller Vn- aller aller Vmn- Min VER:infi pos INSIDE\_narr  
 chez S Sp chez chez chez chez chez Sp Min PRP pos INSIDE\_narr  
 sa D Ds D3fs sa Ds sa sa D- Ds3fss- Min DET:POS pos INSIDE\_narr  
 mère-grand N Nc Nfs mère-grand mère-grand mère-grand mère-grand mère-grand Ncfs Min NOM pos INSIDE\_narr  
 , F , , , , , , F Min PUN punctuation INSIDE\_narr  
 qui P Pr P-ms qui P- qui P- Pr-ms- Min PRO:REL pos INSIDE\_narr  
 demeurait V Vm V3-s demeurait demeurait Vii demeurait demeurait Vmii3s- Min VER:impf pos INSIDE\_narr  
 dans S Sp dans dans dans dans dans dans Sp Min PRP pos INSIDE\_narr  
 un D Da D-ms un D- un un Di Da-ms-i Min DET:ART pos INSIDE\_narr  
 autre A Ai Ams A- autre autre autre autre Ai-ms Min ADJ pos INSIDE\_narr  
 village N Nc Nms village village village village village Ncms Min NOM pos INSIDE\_narr  
 . F . . . . . F Min SENT punctuation INSIDE\_narr  
 \n - - - - - Min - punctuation INSIDE\_narr  
 en S Sp en en en en en Sp Min PRO:PER pos INSIDE\_narr  
 passant V Vm V- passant passant Vpp passant passant Vmpp- Min VER:ppre pos INSIDE\_narr  
 dans S Sp dans dans dans dans dans dans Sp Min PRP pos INSIDE\_narr  
 un D Da D-ms un D- un un Di Da-ms-i Min DET:ART pos INSIDE\_narr  
 bois N Nc Nms bois bois bois bois bois Ncms Min NOM pos INSIDE\_narr  
 elle P Pp P3fs elle P- elle Pn elle Pp3fsn- Min PRO:PER pos INSIDE\_narr  
 rencontra V Vm V3-s rencontra rencontra Vis rencontra rencontra Vmis3s- Min VER:simp pos INSIDE\_narr  
 compère N Nc Nms compère compère compère compère compère Ncms Min NOM pos INSIDE\_narr  
 le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos INSIDE\_narr  
 loup N Nc Nms loup loup loup loup loup Ncms Min NOM pos INSIDE\_narr  
 , F , , , , , , F Min PUN punctuation INSIDE\_narr  
 qui P Pr P-ms qui P- qui P- Pr-ms- Min PRO:REL pos INSIDE\_narr  
 eut V Vm V3-s eut eut Vis eut eut Vmis3s- Min VER:simp pos INSIDE\_narr  
 bien R Rg Rp bien bien bien bien bien Rgp Min ADV pos INSIDE\_narr  
 envie N Nc Nfs envie envie envie envie envie Ncfs Min NOM pos INSIDE\_narr  
 de S Sp de de de de de de de Sp Min PRP pos INSIDE\_narr  
 la P Pp P3fs la P- la Pa la Pp3fsa- Min DET:ART pos INSIDE\_narr  
 manger V Vm V- manger manger Vn- manger manger Vmn- Min NOM pos INSIDE\_narr  
 , F , , , , , , F Min PUN punctuation INSIDE\_narr  
 mais C Cc mais mais mais mais mais mais Cc Min KON pos INSIDE\_narr  
 il P Pp P3ms il P- il Pn il Pp3msn- Min PRO:PER pos INSIDE\_narr  
 n' R Rp Rn n' n' n' n' n' Rpn Min ADV pos INSIDE\_narr  
 osa V Vm V3-s osa osa Vis osa osa Vmis3s- Min VER:simp pos INSIDE\_narr  
 , F , , , , , , F Min PUN punctuation INSIDE\_narr  
 à S Sp à à à à à Sp Min PRP pos INSIDE\_narr  
 cause N Nc Nfs cause cause cause cause cause Ncfs Min NOM pos INSIDE\_narr  
 de S Sp de de de de de de de Sp Min PRP pos INSIDE\_narr  
 quelques D Di D-mp quelques D- quelques quelques D- Di-mp- Min PRO:IND pos INSIDE\_narr  
 bûcherons N Nc Nmp bûcherons bûcherons bûcherons bûcherons Ncmp Min NOM pos INSIDE\_narr  
 qui P Pr P-mp qui P- qui P- Pr-mp- Min PRO:REL pos INSIDE\_narr  
 étaient V Vm V3-p étaient étaient Vii étaient étaient Vmii3p- Min VER:impf pos INSIDE\_narr  
 dans S Sp dans dans dans dans dans dans Sp Min PRP pos INSIDE\_narr  
 la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos INSIDE\_narr  
 forêt N Nc Nfs forêt forêt forêt forêt forêt Ncfs Min NOM pos INSIDE\_narr  
 . F . . . . . F Min SENT punctuation INSIDE\_narr  
 \n - - - - - Min - punctuation INSIDE\_narr  
 il P Pp P3ms il P- il Pn il Pp3msn- Min PRO:PER pos INSIDE\_narr  
 lui P Pp P3ms lui P- lui Pd lui Pp3msd- Min PRO:PER pos INSIDE\_narr  
 demanda V Vm V3-s demanda demanda Vis demanda demanda Vmis3s- Min VER:simp pos INSIDE\_narr  
 où P Pr P-ms où P- où P- Pr-ms- Min PRO:REL pos INSIDE\_narr  
 elle P Pp P3fs elle P- elle Pn elle Pp3fsn- Min PRO:PER pos INSIDE\_narr  
 allait V Vm V3-s allait allait Vii allait allait Vmii3s- Min VER:impf pos INSIDE\_narr  
 ; F ; ; ; ; ; F Min PUN punctuation INSIDE\_narr  
 la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos INSIDE\_narr



pauvre A Af Afs Ap pauvre pauvre pauvre pauvre Afpfs Min ADJ pos INSIDE\_narr  
enfant N Nc Nfs enfant enfant enfant enfant enfant Ncfs Min NOM pos INSIDE\_narr  
, F , , , , , F Min PUN punctuation INSIDE\_narr  
qui P Pr P-ms qui P- qui P- qui Pr-ms— Min PRO:REL pos INSIDE\_narr  
ne R Rp Rn ne ne ne ne ne Rpn Min ADV pos INSIDE\_narr  
savait V Vm V3-s savait savait Vii savait savait Vmii3s- Min VER:impf pos INSIDE\_narr  
pas R Rg Rn pas pas pas pas Rgn Min ADV pos INSIDE\_narr  
qu' C Cs qu' qu' qu' qu' qu' qu' Cs Min KON pos INSIDE\_narr  
il P Pp P3ms il P- il Pn il Pp3msn- Min PRO:PER pos INSIDE\_narr  
est V Vm V3-s est est Vip est est Vmip3s- Min VER:pres pos INSIDE\_narr  
dangereux A Af Ams Ap dangereux dangereux dangereux dangereux Afpms Min ADJ pos INSIDE\_narr  
de S Sp de de de de de de Sp Min PRP pos INSIDE\_narr  
s' P Px P3ms s' P- s' P- s' Px3ms— Min PRO:PER pos INSIDE\_narr  
arrêter V Vm V— arrêter arrêter Vn- arrêter arrêter Vmn— Min VER:infi pos INSIDE\_narr  
à S Sp à à à à à à Sp Min PRP pos INSIDE\_narr  
écouter V Vm V— écouter écouter Vn- écouter écouter Vmn— Min VER:infi pos INSIDE\_narr  
un D Da D-ms un D- un un Di Da-ms-i Min DET:ART pos INSIDE\_narr  
loup N Nc Nms loup loup loup loup loup loup Ncms Min NOM pos INSIDE\_narr  
, F , , , , , F Min PUN punctuation INSIDE\_narr  
lui P Pp P3ms lui P- lui Pd lui Pp3msd- Min PRO:PER pos INSIDE\_narr  
dit V Vm V3-s dit dit Vip dit dit Vmip3s- Min VER:pres pos INSIDE\_narr  
: F : : : : : F Min PUN punctuation INSIDE\_narr  
\n - - - - - Min - punctuation INSIDE\_narr  
- F - - - - - F Min PUN punctuation BEGIN\_spkr  
" F " " " " " " F Min PUN: cit punctuation INSIDE\_spkr  
je P Pp Plms je P- je Pn je Pplmsn- Min PRO:PER pos INSIDE\_spkr  
vais V Vm V1-s vais vais Vip vais vais Vmip1s- Min VER:pres pos INSIDE\_spkr  
voir V Vm V— voir voir Vn- voir voir Vmn— Min VER:infi pos INSIDE\_spkr  
ma D Ds D1fs ma Ds ma ma D- Ds1fss- Min DET:POS pos INSIDE\_spkr  
mère-grand N Nc Nfs mère-grand mère-grand mère-grand mère-grand Ncfs Min NOM pos INSIDE\_spkr  
, F , , , , , F Min PUN punctuation INSIDE\_spkr  
et C Cc et et et et et Cc Min KON pos INSIDE\_spkr  
lui P Pp P3ms lui P- lui Pd lui Pp3msd- Min PRO:PER pos INSIDE\_spkr  
porter V Vm V— porter porter Vn- porter porter Vmn— Min VER:infi pos INSIDE\_spkr  
une D Da D-fs une D- une une Di Da-fs-i Min DET:ART pos INSIDE\_spkr  
galette N Nc Nfs galette galette galette galette galette Ncfs Min NOM pos INSIDE\_spkr  
avec S Sp avec avec avec avec avec Sp Min PRP pos INSIDE\_spkr  
un D Da D-ms un D- un un Di Da-ms-i Min DET:ART pos INSIDE\_spkr  
petit A Af Ams Ap petit petit petit petit Afpms Min ADJ pos INSIDE\_spkr  
pot N Nc Nms pot pot pot pot pot Ncms Min NOM pos INSIDE\_spkr  
de S Sp de de de de de de Sp Min PRP pos INSIDE\_spkr  
beurre N Nc Nms beurre beurre beurre beurre Ncms Min NOM pos INSIDE\_spkr  
que C Cs que que que que que Cs Min KON pos INSIDE\_spkr  
ma D Ds D1fs ma Ds ma ma D- Ds1fss- Min DET:POS pos INSIDE\_spkr  
mère N Nc Nfs mère mère mère mère Ncfs Min NOM pos INSIDE\_spkr  
lui P Pp P3ms lui P- lui Pd lui Pp3msd- Min PRO:PER pos INSIDE\_spkr  
envoie V Vm V3-s envoie envoie Vip envoie envoie Vmip3s- Min VER:pres pos INSIDE\_spkr  
. F . . . . . F Min SENT punctuation INSIDE\_spkr  
" F " " " " " " F Min PUN: cit punctuation INSIDE\_spkr  
\n - - - - - Min - punctuation INSIDE\_spkr  
- F - - - - - F Min PUN punctuation BEGIN\_spkr  
" F " " " " " " F Min PUN: cit punctuation INSIDE\_spkr  
demeure V Vm V3-s demeure demeure Vip demeure demeure Vmip3s- Min VER:pres pos INSIDE\_spkr  
-t-elle P Pp P3fs -t-elle P- -t-elle Pn -t-elle Pp3fsn- Min PRO:PER pos INSIDE\_spkr  
bien R Rg Rp bien bien bien bien bien Rgp Min ADV pos INSIDE\_spkr  
loin R Rg Rp loin loin loin loin loin Rgp Min ADV pos INSIDE\_spkr  
? F ? ? ? ? ? F Min SENT punctuation INSIDE\_spkr  
" F " " " " " " F Min PUN: cit punctuation INSIDE\_spkr  
lui P Pp P3ms lui P- lui Pd lui Pp3msd- Min PRO:PER pos BEGIN\_narr  
dit V Vm V3-s dit dit Vip dit dit Vmip3s- Min VER:pres pos INSIDE\_narr  
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos INSIDE\_narr  
loup N Nc Nms loup loup loup loup loup Ncms Min NOM pos INSIDE\_narr  
. F . . . . . F Min SENT punctuation INSIDE\_narr  
\n - - - - - Min - punctuation INSIDE\_narr  
- F - - - - - F Min PUN punctuation BEGIN\_spkr  
" F " " " " " " F Min PUN: cit punctuation INSIDE\_spkr  
oh I oh oh oh oh oh oh oh I Min INT pos INSIDE\_spkr  
! F ! ! ! ! ! F Min SENT punctuation INSIDE\_spkr  
oui R Rg Rp oui oui oui oui oui Rgp Min INT pos INSIDE\_spkr  
" F " " " " " " F Min PUN: cit punctuation INSIDE\_spkr  
, F , , , , , F Min PUN punctuation INSIDE\_spkr  
dit V Vm V3-s dit dit Vip dit dit Vmip3s- Min VER:pres pos BEGIN\_narr  
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos INSIDE\_narr  
petit A Af Ams Ap petit petit petit petit Afpms Min ADJ pos INSIDE\_narr  
chaperon N Nc Nms chaperon chaperon chaperon chaperon Ncms Min NOM pos INSIDE\_narr  
rouge A Af Ams Ap rouge rouge rouge rouge Afpms Min ADJ pos INSIDE\_narr  
, F , , , , , F Min PUN punctuation INSIDE\_narr  
" F " " " " " " F Min PUN: cit punctuation BEGIN\_spkr

c' P Pd P-ms c' P- c' P- c' Pd-ms— Min PRO:DEM pos INSIDE\_spkr  
 est V Vm V3-s est est Vip est est Vmip3s- Min VER:pres pos INSIDE\_spkr  
 par S Sp par par par par par Sp Min PRP pos INSIDE\_spkr  
 delà R Rg Rp delà delà delà delà delà Rgp Min ADV pos INSIDE\_spkr  
 le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos INSIDE\_spkr  
 moulin N Nc Nms moulin moulin moulin moulin Ncms Min NOM pos INSIDE\_spkr  
 que C Cs que que que que que Cs Min PRO:REL pos INSIDE\_spkr  
 vous P Pp P2mp vous P- vous Pn vous Pp2mpn- Min PRO:PER pos INSIDE\_spkr  
 voyez V Vm V2-p voyez voyez Vip voyez voyez Vmip2p- Min VER:pres pos INSIDE\_spkr  
 tout R Rg Rp tout tout tout tout tout Rgp Min ADV pos INSIDE\_spkr  
 là-bas R Rg Rp là-bas là-bas là-bas là-bas là-bas Rgp Min ADV pos INSIDE\_spkr  
 , F , , , , , , F Min PUN punctuation INSIDE\_spkr  
 là-bas R Rg Rp là-bas là-bas là-bas là-bas là-bas Rgp Min ADV pos INSIDE\_spkr  
 , F , , , , , , F Min PUN punctuation INSIDE\_spkr  
 à S Sp à à à à à à Sp Min PRP pos INSIDE\_spkr  
 la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos INSIDE\_spkr  
 première A Ao Afs A- première première première première Ao-fs Min NUM pos INSIDE\_spkr  
 maison N Nc Nfs maison maison maison maison maison Ncfs Min NOM pos INSIDE\_spkr  
 du S+D Sp+Da du+D-ms du+du du+D- du+du du+du du+Dd Sp+Da-ms-d Min PRP:det pos INSIDE\_spkr  
 village N Nc Nms village village village village village Ncms Min NOM pos INSIDE\_spkr  
 . F . . . . . F Min SENT punctuation INSIDE\_spkr  
 " F " " " " " " F Min PUN:cit punctuation INSIDE\_spkr  
 \n - - - - - Min - punctuation INSIDE\_spkr  
 - F - - - - - F Min PUN punctuation BEGIN\_spkr  
 " F " " " " " " F Min PUN:cit punctuation INSIDE\_spkr  
 eh I eh eh eh eh eh eh I Min INT pos INSIDE\_spkr  
 bien R Rg Rp bien bien bien bien bien Rgp Min ADV pos INSIDE\_spkr  
 ! F ! ! ! ! ! ! F Min SENT punctuation INSIDE\_spkr  
 " F " " " " " " F Min PUN:cit punctuation INSIDE\_spkr  
 , F , , , , , , F Min PUN punctuation INSIDE\_spkr  
 dit V Vm V3-s dit dit Vip dit dit Vmip3s- Min VER:pres pos BEGIN\_narr  
 le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos INSIDE\_narr  
 loup N Nc Nms loup loup loup loup loup Ncms Min NOM pos INSIDE\_narr  
 , F , , , , , , F Min PUN punctuation INSIDE\_narr  
 " F " " " " " " F Min PUN:cit punctuation BEGIN\_spkr  
 je P Pp Plms je P- je Pn je Pplmsn- Min PRO:PER pos INSIDE\_spkr  
 veux V Vm V1-s veux veux Vip veux veux Vmip1s- Min VER:pres pos INSIDE\_spkr  
 y P Pp P3ms y P- y Pd y Pp3msd- Min PRO:PER pos INSIDE\_spkr  
 aller V Vm V- aller aller Vn- aller aller Vmn- Min VER:infi pos INSIDE\_spkr  
 voir V Vm V- voir voir Vn- voir voir Vmn- Min VER:infi pos INSIDE\_spkr  
 aussi R Rg Rp aussi aussi aussi aussi aussi Rgp Min ADV pos INSIDE\_spkr  
 ; F ; ; ; ; ; ; F Min PUN punctuation INSIDE\_spkr  
 je P Pp Plms je P- je Pn je Pplmsn- Min PRO:PER pos INSIDE\_spkr  
 m' P Px Plms m' P- m' P- m' Px1ms- Min PRO:PER pos INSIDE\_spkr  
 y P Pp P3ms y P- y Pd y Pp3msd- Min PRO:PER pos INSIDE\_spkr  
 en P Pp P3ms en P- en Pd en Pp3msd- Min PRO:PER pos INSIDE\_spkr  
 vais V Vm V1-s vais vais Vip vais vais Vmip1s- Min VER:pres pos INSIDE\_spkr  
 par S Sp par par par par par par Sp Min PRP pos INSIDE\_spkr  
 ce D Dd D-ms ce D- ce ce D- Dd-ms- Min PRO:DEM pos INSIDE\_spkr  
 chemin N Nc Nms chemin chemin chemin chemin chemin Ncms Min NOM pos INSIDE\_spkr  
 -ci R Rg Rp -ci -ci -ci -ci -ci Rgp Min ADV pos INSIDE\_spkr  
 , F , , , , , , F Min PUN punctuation INSIDE\_spkr  
 et C Cc et et et et et Cc Min KON pos INSIDE\_spkr  
 toi P Pp P2ms toi P- toi Po toi Pp2mso- Min PRO:PER pos INSIDE\_spkr  
 par S Sp par par par par par par Sp Min PRP pos INSIDE\_spkr  
 ce D Dd D-ms ce D- ce ce D- Dd-ms- Min PRO:DEM pos INSIDE\_spkr  
 chemin N Nc Nms chemin chemin chemin chemin chemin Ncms Min NOM pos INSIDE\_spkr  
 -là R Rg Rp -là -là -là -là -là Rgp Min ADV pos INSIDE\_spkr  
 , F , , , , , , F Min PUN punctuation INSIDE\_spkr  
 et C Cc et et et et et Cc Min KON pos INSIDE\_spkr  
 nous P Pp Plmp nous P- nous Pn nous Pplmpn- Min PRO:PER pos INSIDE\_spkr  
 verrons V Vm V1-p verrons verrons Vif verrons verrons Vmiflp- Min VER:futu pos INSIDE\_spkr  
 qui P Pr P-mp qui P- qui P- qui Pr-mp- Min PRO:REL pos INSIDE\_spkr  
 plus R Rg Rc plus plus plus plus plus Rgc Min ADV pos INSIDE\_spkr  
 tôt R Rg Rp tôt tôt tôt tôt tôt Rgp Min ADV pos INSIDE\_spkr  
 y P Pp P3ms y P- y Pd y Pp3msd- Min PRO:PER pos INSIDE\_spkr  
 sera V Vm V3-s sera sera Vif sera sera Vmif3s- Min VER:futu pos INSIDE\_spkr  
 . F . . . . . F Min SENT punctuation INSIDE\_spkr  
 " F " " " " " " F Min PUN:cit punctuation INSIDE\_spkr  
 \n \n - - - - - Min - punctuation INSIDE\_spkr  
 le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos BEGIN\_narr  
 loup N Nc Nms loup loup loup loup loup Ncms Min NOM pos INSIDE\_narr  
 se P Px P3ms se P- se P- se Px3ms- Min PRO:PER pos INSIDE\_narr  
 mit V Vm V3-s mit mit Vis mit mit Vmif3s- Min VER:simp pos INSIDE\_narr  
 à S Sp à à à à à à Sp Min PRP pos INSIDE\_narr  
 courir V Vm V- courir courir Vn- courir courir Vmn- Min VER:infi pos INSIDE\_narr  
 de S Sp de de de de de de Sp Min PRP pos INSIDE\_narr  
 toute D Di D-fs toute D- toute toute D- Di-fs- Min PRO:IND pos INSIDE\_narr

sa D Ds D3fs sa Ds sa sa D- Ds3fss- Min DET:POS pos INSIDE\_narr  
force N Nc Nfs force force force force force Ncfs Min NOM pos INSIDE\_narr  
par S Sp par par par par par Sp Min PRP pos INSIDE\_narr  
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos INSIDE\_narr  
chemin N Nc Nms chemin chemin chemin chemin chemin Ncms Min NOM pos INSIDE\_narr  
qui P Pr P-ms qui P- qui P- qui Pr-ms- Min PRO:REL pos INSIDE\_narr  
était V Vm V3-s était était Vii était était Vmii3s- Min VER:impf pos INSIDE\_narr  
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos INSIDE\_narr  
plus R Rg Rc plus plus plus plus plus Rgc Min ADV pos INSIDE\_narr  
court A Af Ams Ap court court court court court Afpms Min ADJ pos INSIDE\_narr  
, F , , , , , F Min PUN punctuation INSIDE\_narr  
et C Cc et et et et et Cc Min KON pos INSIDE\_narr  
la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos INSIDE\_narr  
petite A Af Afs Ap petite petite petite petite Afpfs Min ADJ pos INSIDE\_narr  
fille N Nc Nfs fille fille fille fille fille Ncfs Min NOM pos INSIDE\_narr  
s' P Px P3fs s' P- s' P- s' Px3fs- Min PRO:PER pos INSIDE\_narr  
en P Pp P3ms en P- en Pd en Pp3msd- Min PRO:PER pos INSIDE\_narr  
alla V Vm V3-s alla alla Vis alla alla Vmis3s- Min VER:simp pos INSIDE\_narr  
par S Sp par par par par par par Sp Min PRP pos INSIDE\_narr  
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos INSIDE\_narr  
chemin N Nc Nms chemin chemin chemin chemin chemin Ncms Min NOM pos INSIDE\_narr  
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos INSIDE\_narr  
plus R Rg Rc plus plus plus plus plus Rgc Min ADV pos INSIDE\_narr  
long A Af Ams Ap long long long long long Afpms Min ADJ pos INSIDE\_narr  
, F , , , , , F Min PUN punctuation INSIDE\_narr  
s' P Px P3ms s' P- s' P- s' Px3ms- Min PRO:PER pos INSIDE\_narr  
amusant V Vm V- amusant amusant Vpp amusant amusant Vmpp- Min VER:ppre pos INSIDE\_narr  
à S Sp à à à à à à Sp Min PRP pos INSIDE\_narr  
cueillir V Vm V- cueillir cueillir Vn- cueillir cueillir Vmn- Min VER:infi pos INSIDE\_narr  
des D Da D-fp des D- des des Di Da-fp-i Min PRP:det pos INSIDE\_narr  
noisettes N Nc Nfp noisettes noisettes noisettes noisettes noisettes Ncfp Min NOM pos INSIDE\_narr  
, F , , , , , F Min PUN punctuation INSIDE\_narr  
à S Sp à à à à à à Sp Min PRP pos INSIDE\_narr  
courir V Vm V- courir courir Vn- courir courir Vmn- Min VER:infi pos INSIDE\_narr  
après S Sp après après après après après Sp Min PRP pos INSIDE\_narr  
des D Da D-mp des D- des des Di Da-mp-i Min PRP:det pos INSIDE\_narr  
papillons N Nc Nmp papillons papillons papillons papillons papillons Ncmp Min NOM pos INSIDE\_narr  
, F , , , , , F Min PUN punctuation INSIDE\_narr  
et C Cc et et et et et Cc Min KON pos INSIDE\_narr  
à S Sp à à à à à à Sp Min PRP pos INSIDE\_narr  
faire V Vm V- faire faire Vn- faire faire Vmn- Min VER:infi pos INSIDE\_narr  
des D Da D-mp des D- des des Di Da-mp-i Min PRP:det pos INSIDE\_narr  
bouquets N Nc Nmp bouquets bouquets bouquets bouquets bouquets Ncmp Min NOM pos INSIDE\_narr  
des S+D Sp+Da des+D-fp des+des des+D- des+des des+des des+Dd Sp+Da-fp-d Min PRP:det pos INSIDE\_narr  
petites A Af Afp Ap petites petites petites petites Afpfp Min ADJ pos INSIDE\_narr  
fleurs N Nc Nfp fleurs fleurs fleurs fleurs fleurs Ncfp Min NOM pos INSIDE\_narr  
qu' P Pr P-fp qu' P- qu' P- qu' Pr-fp- Min PRO:REL pos INSIDE\_narr  
elle P Pp P3fs elle P- elle Pn elle Pp3fsn- Min PRO:PER pos INSIDE\_narr  
rencontrait V Vm V3-s rencontrait rencontrait Vii rencontrait rencontrait Vmii3s- Min VER:impf pos INSIDE\_narr  
. F . . . . . F Min SENT punctuation INSIDE\_narr  
\n\n - - - - - Min - punctuation INSIDE\_narr  
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos INSIDE\_narr  
loup N Nc Nms loup loup loup loup loup Ncms Min NOM pos INSIDE\_narr  
ne R Rp Rn ne ne ne ne ne Rpn Min ADV pos INSIDE\_narr  
fut V Va V3-s fut fut Vis fut fut Vais3s- Min VER:simp pos INSIDE\_narr  
pas R Rg Rn pas pas pas pas pas Rgn Min ADV pos INSIDE\_narr  
longtemps R Rg Rp longtemps longtemps longtemps longtemps longtemps Rgp Min ADV pos INSIDE\_narr  
à S Sp à à à à à à Sp Min PRP pos INSIDE\_narr  
arriver V Vm V- arriver arriver Vn- arriver arriver Vmn- Min VER:infi pos INSIDE\_narr  
à S Sp à à à à à à Sp Min PRP pos INSIDE\_narr  
la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos INSIDE\_narr  
maison N Nc Nfs maison maison maison maison maison Ncfs Min NOM pos INSIDE\_narr  
de S Sp de de de de de de Sp Min PRP pos INSIDE\_narr  
la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos INSIDE\_narr  
mère-grand N Nc Nfs mère-grand mère-grand mère-grand mère-grand mère-grand Ncfs Min NOM pos INSIDE\_narr  
; F ; ; ; ; ; F Min PUN punctuation INSIDE\_narr  
il P Pp P3ms il P- il Pn il Pp3msn- Min PRO:PER pos INSIDE\_narr  
heurte V Vm V3-s heurte heurte Vip heurte heurte Vmip3s- Min VER:pres pos INSIDE\_narr  
: F : : : : : F Min PUN punctuation INSIDE\_narr  
toc N Np Nms toc toc toc toc toc Npms Min ADJ pos INSIDE\_narr  
, F , , , , , F Min PUN punctuation INSIDE\_narr  
toc N Np Nms toc toc toc toc toc Npms Min INT pos INSIDE\_narr  
. F . . . . . F Min SENT punctuation INSIDE\_narr  
\n\n - - - - - Min - punctuation INSIDE\_narr  
- F - - - - - F Min PUN punctuation BEGIN\_spkr  
" F " " " " " " F Min PUN: cit punctuation INSIDE\_spkr  
qui P Pr P-ms qui P- qui P- qui Pr-ms- Min PRO:REL pos INSIDE\_spkr  
est V Vm V3-s est est Vip est est Vmip3s- Min VER:pres pos INSIDE\_spkr

## C.1. Données d'entrée

```

là R Rg Rp là là là là là Rgp Min ADV pos      INSIDE_spkr
? F ? ? ? ? ? ? ? F Min SENT punctuation     INSIDE_spkr
" F " " " " " " " F Min PUN:cit punctuation  INSIDE_spkr
\n - - - - - Min - punctuation INSIDE_spkr
- F - - - - - F Min PUN punctuation BEGIN_spkr
" F " " " " " " " F Min PUN:cit punctuation  INSIDE_spkr
c' P Pd P-ms c' P- c' P- c' Pd-ms- Min PRO:DEM pos      INSIDE_spkr
est V Vm V3-s est est Vip est est Vmip3s- Min VER:pres pos      INSIDE_spkr
votre D Ds D2fs votre Dp votre votre D- Ds2fsp- Min DET:POS pos      INSIDE_spkr
fille N Nc Nfs fille fille fille fille Ncfs Min NOM pos      INSIDE_spkr
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos      INSIDE_spkr
petit A Af Ams Ap petit petit petit petit Afpms Min ADJ pos      INSIDE_spkr
chaperon N Nc Nms chaperon chaperon chaperon chaperon chaperon Ncms Min NOM pos      INSIDE_spkr
rouge A Af Ams Ap rouge rouge rouge rouge Afpms Min ADJ pos      INSIDE_spkr
" F " " " " " " " F Min PUN:cit punctuation  INSIDE_spkr
( F ( ( ( ( ( ( F Min PUN punctuation BEGIN_narr
dit V Vm V3-s dit dit Vip dit dit Vmip3s- Min VER:pres pos      INSIDE_narr
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos      INSIDE_narr
loup N Nc Nms loup loup loup loup loup loup Ncms Min NOM pos      INSIDE_narr
, F , , , , , , F Min PUN punctuation INSIDE_narr
en S Sp en en en en en Sp Min PRO:PER pos      INSIDE_narr
contrefaisant V Vm V- contrefaisant contrefaisant Vn- contrefaisant contrefaisant Vmn- Min VER:ppre pos      INSIDE_narr
sa D Ds D3fs sa Ds sa sa D- Ds3fss- Min DET:POS pos      INSIDE_narr
voix N Nc Nfs voix voix voix voix voix Ncfs Min NOM pos      INSIDE_narr
) F ) ) ) ) ) ) F Min PUN punctuation INSIDE_narr
" F " " " " " " " F Min PUN:cit punctuation  BEGIN_spkr
qui P Pr P-ms qui P- qui P- qui Pr-ms- Min PRO:REL pos      INSIDE_spkr
vous P Pp P2mp vous P- vous Pa vous Pp2mpa- Min PRO:PER pos      INSIDE_spkr
apporte V Vm V3-s apporte apporte Vip apporte apporte Vmip3s- Min VER:pres pos      INSIDE_spkr
une D Da D-fs une D- une une Di Da-fs-i Min DET:ART pos      INSIDE_spkr
galette N Nc Nfs galette galette galette galette galette Ncfs Min NOM pos      INSIDE_spkr
et C Cc et et et et et Cc Min KON pos      INSIDE_spkr
un D Da D-ms un D- un un Di Da-ms-i Min DET:ART pos      INSIDE_spkr
petit A Af Ams Ap petit petit petit petit Afpms Min ADJ pos      INSIDE_spkr
pot N Nc Nms pot pot pot pot pot Ncms Min NOM pos      INSIDE_spkr
de S Sp de de de de de de Sp Min PRP pos      INSIDE_spkr
beurre N Nc Nms beurre beurre beurre beurre beurre Ncms Min NOM pos      INSIDE_spkr
que C Cs que que que que que Cs Min KON pos      INSIDE_spkr
ma D Ds D1fs ma Ds ma ma D- Ds1fss- Min DET:POS pos      INSIDE_spkr
mère N Nc Nfs mère mère mère mère mère Ncfs Min NOM pos      INSIDE_spkr
vous P Pp P2mp vous P- vous Pa vous Pp2mpa- Min PRO:PER pos      INSIDE_spkr
envoie V Vm V1-s envoie envoie Vip envoie envoie Vmip1s- Min VER:pres pos      INSIDE_spkr
. F . . . . . F Min SENT punctuation      INSIDE_spkr
" F " " " " " " " F Min PUN:cit punctuation  INSIDE_spkr
\n - - - - - Min - punctuation INSIDE_spkr
la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos      BEGIN_narr
bonne A Af Afs Ap bonne bonne bonne bonne Afpfs Min ADJ pos      INSIDE_narr
mère-grand N Nc Nfs mère-grand mère-grand mère-grand mère-grand mère-grand Ncfs Min NOM pos      INSIDE_narr
, F , , , , , , F Min PUN punctuation INSIDE_narr
qui P Pr P-ms qui P- qui P- qui Pr-ms- Min PRO:REL pos      INSIDE_narr
était V Vm V3-s était était Vii était était Vmii3s- Min VER:impf pos      INSIDE_narr
dans S Sp dans dans dans dans dans dans Sp Min PRP pos      INSIDE_narr
son D Ds D3ms son Ds son son D- Ds3mss- Min DET:POS pos      INSIDE_narr
lit N Nc Nms lit lit lit lit lit Ncms Min NOM pos      INSIDE_narr
, F , , , , , , F Min PUN punctuation INSIDE_narr
car C Cc car car car car car Cc Min KON pos      INSIDE_narr
elle P Pp P3fs elle P- elle Pn elle Pp3fsn- Min PRO:PER pos      INSIDE_narr
se P Px P3fs se P- se P- se Px3fs- Min PRO:PER pos      INSIDE_narr
trouvait V Vm V3-s trouvait trouvait Vii trouvait trouvait Vmii3s- Min VER:impf pos      INSIDE_narr
un D Da D-ms un D- un un Di Da-ms-i Min DET:ART pos      INSIDE_narr
peu R Rg Rp peu peu peu peu peu peu Rgp Min ADV pos      INSIDE_narr
mal R Rg Rp mal mal mal mal mal mal Rgp Min ADJ pos      INSIDE_narr
, F , , , , , , F Min PUN punctuation INSIDE_narr
lui P Pp P3ms lui P- lui Pd lui Pp3msd- Min PRO:PER pos      INSIDE_narr
cria V Vm V3-s cria cria Vis cria cria Vmis3s- Min VER:simp pos      INSIDE_narr
: F : : : : : : F Min PUN punctuation INSIDE_narr
\n - - - - - Min - punctuation INSIDE_narr
- F - - - - - F Min PUN punctuation BEGIN_spkr
" F " " " " " " " F Min PUN:cit punctuation  INSIDE_spkr
tire V Vm V3-s tire tire Vip tire tire Vmip3s- Min VER:pres pos      INSIDE_spkr
la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos      INSIDE_spkr
chevillette N Nc Nfs chevillette chevillette chevillette chevillette chevillette Ncfs Min NOM pos      INSIDE_spkr
, F , , , , , , F Min PUN punctuation INSIDE_spkr
la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos      INSIDE_spkr
bobinette N Nc Nfs bobinette bobinette bobinette bobinette bobinette Ncfs Min NOM pos      INSIDE_spkr
cherra A Af Afs Ap cherra cherra cherra cherra cherra Afpfs Min VER:simp pos      INSIDE_spkr
. F . . . . . F Min SENT punctuation      INSIDE_spkr
" F " " " " " " " F Min PUN:cit punctuation  INSIDE_spkr

```

\n\n - - - - - Min - punctuation INSIDE\_spkr  
 le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos BEGIN\_narr  
 loup N Nc Nms loup loup loup loup loup Ncms Min NOM pos INSIDE\_narr  
 tira V Vm V3-s tira tira Vis tira tira Vmis3s- Min VER:simp pos INSIDE\_narr  
 la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos INSIDE\_narr  
 chevillette N Nc Nfs chevillette chevillette chevillette chevillette chevillette Ncfs Min NOM pos INSIDE\_narr  
 , F , , , , , F Min PUN punctuation INSIDE\_narr  
 et C Cc et et et et et et Cc Min KON pos INSIDE\_narr  
 la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos INSIDE\_narr  
 porte N Nc Nfs porte porte porte porte porte Ncfs Min NOM pos INSIDE\_narr  
 s' P Px P3fs s' P- s' P- s' Px3fs- Min PRO:PER pos INSIDE\_narr  
 ouvrit V Vm V3-s ouvrit ouvrit Vis ouvrit ouvrit Vmis3s- Min VER:simp pos INSIDE\_narr  
 . F . . . . . F Min SENT punctuation INSIDE\_narr  
 \n - - - - - Min - punctuation INSIDE\_narr  
 il P Pp P3ms il P- il Pn il Pp3msn- Min PRO:PER pos INSIDE\_narr  
 se P Px P3ms se P- se P- se Px3ms- Min PRO:PER pos INSIDE\_narr  
 jeta V Vm V3-s jeta jeta Vis jeta jeta Vmis3s- Min VER:simp pos INSIDE\_narr  
 sur S Sp sur sur sur sur sur Sp Min PRP pos INSIDE\_narr  
 la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos INSIDE\_narr  
 bonne A Af Afs Ap bonne bonne bonne bonne Afpfs Min ADJ pos INSIDE\_narr  
 femme N Nc Nfs femme femme femme femme Ncfs Min NOM pos INSIDE\_narr  
 , F , , , , , F Min PUN punctuation INSIDE\_narr  
 et C Cc et et et et et et Cc Min KON pos INSIDE\_narr  
 la P Pp P3fs la P- la Pa la Pp3fsa- Min PRO:PER pos INSIDE\_narr  
 dévora V Vm V3-s dévora dévora Vis dévora dévora Vmis3s- Min VER:simp pos INSIDE\_narr  
 en S Sp en en en en en en Sp Min PRP pos INSIDE\_narr  
 moins N Nc Nms moins moins moins moins moins Ncms Min ADV pos INSIDE\_narr  
 de S Sp de de de de de de Sp Min PRP pos INSIDE\_narr  
 rien P Pi P-ms rien P- rien P- rien Pi-ms- Min NOM pos INSIDE\_narr  
 ; F ; ; ; ; ; F Min PUN punctuation INSIDE\_narr  
 car C Cc car car car car Cc Min KON pos INSIDE\_narr  
 il P Pp P3ms il P- il Pn il Pp3msn- Min PRO:PER pos INSIDE\_narr  
 y P Pp P3ms y P- y Pd y Pp3msd- Min PRO:PER pos INSIDE\_narr  
 avait V Va V3-s avait avait Vii avait avait Vaii3s- Min VER:impf pos INSIDE\_narr  
 plus R Rg Rn plus plus plus plus Rgn Min ADV pos INSIDE\_narr  
 de S Sp de de de de de de Sp Min PRP pos INSIDE\_narr  
 trois D Dk D-mp trois D- trois trois D- Dk-mp- Min NUM pos INSIDE\_narr  
 jours N Nc Nmp jours jours jours jours jours Ncmp Min NOM pos INSIDE\_narr  
 qu' C Cs qu' qu' qu' qu' qu' qu' Cs Min KON pos INSIDE\_narr  
 il P Pp P3ms il P- il Pn il Pp3msn- Min PRO:PER pos INSIDE\_narr  
 n' R Rp Rn n' n' n' n' n' n' Rpn Min ADV pos INSIDE\_narr  
 avait V Va V3-s avait avait Vii avait avait Vaii3s- Min VER:impf pos INSIDE\_narr  
 mangé V Vm V-ms mangé mangé Vps mangé mangé Vmps-sm Min VER:pper pos INSIDE\_narr  
 . F . . . . . F Min SENT punctuation INSIDE\_narr  
 \n - - - - - Min - punctuation INSIDE\_narr  
 ensuite R Rg Rp ensuite ensuite ensuite ensuite Rgp Min ADV pos INSIDE\_narr  
 il P Pp P3ms il P- il Pn il Pp3msn- Min PRO:PER pos INSIDE\_narr  
 ferma V Vm V3-s ferma ferma Vis ferma ferma Vmis3s- Min VER:simp pos INSIDE\_narr  
 la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos INSIDE\_narr  
 porte N Nc Nfs porte porte porte porte porte Ncfs Min NOM pos INSIDE\_narr  
 , F , , , , , F Min PUN punctuation INSIDE\_narr  
 et C Cc et et et et et et Cc Min KON pos INSIDE\_narr  
 s' P Px P3ms s' P- s' P- s' Px3ms- Min PRO:PER pos INSIDE\_narr  
 alla V Vm V3-s alla alla Vis alla alla Vmis3s- Min VER:simp pos INSIDE\_narr  
 coucher V Vm V- coucher coucher Vn- coucher coucher Vmn- Min VER:infi pos INSIDE\_narr  
 dans S Sp dans dans dans dans dans Sp Min PRP pos INSIDE\_narr  
 le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos INSIDE\_narr  
 lit N Nc Nms lit lit lit lit lit Ncms Min NOM pos INSIDE\_narr  
 de S Sp de de de de de de Sp Min PRP pos INSIDE\_narr  
 la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos INSIDE\_narr  
 mère-grand N Nc Nfs mère-grand mère-grand mère-grand mère-grand mère-grand Ncfs Min NOM pos INSIDE\_narr  
 , F , , , , , F Min PUN punctuation INSIDE\_narr  
 en S Sp en en en en en en Sp Min PRP pos INSIDE\_narr  
 attendant V Vm V- attendant attendant Vpp attendant attendant Vmpp- Min VER:ppre pos INSIDE\_narr  
 le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos INSIDE\_narr  
 petit A Af Ams Ap petit petit petit petit Afpms Min ADJ pos INSIDE\_narr  
 chaperon N Nc Nms chaperon chaperon chaperon chaperon Ncms Min NOM pos INSIDE\_narr  
 rouge A Af Ams Ap rouge rouge rouge rouge Afpms Min ADJ pos INSIDE\_narr  
 , F , , , , , F Min PUN punctuation INSIDE\_narr  
 qui P Pr P-ms qui P- qui P- qui P-ms- Min PRO:REL pos INSIDE\_narr  
 quelque D Di D-ms quelque D- quelque quelque D- Di-ms- Min PRO:IND pos INSIDE\_narr  
 temps N Nc Nms temps temps temps temps temps Ncms Min NOM pos INSIDE\_narr  
 après S Sp après après après après après Sp Min KON pos INSIDE\_narr  
 vint V Vm V3-s vint vint Vis vint vint Vmis3s- Min VER:simp pos INSIDE\_narr  
 heurter V Vm V- heurter heurter Vn- heurter heurter Vmn- Min VER:infi pos INSIDE\_narr  
 à S Sp à à à à à à Sp Min PRP pos INSIDE\_narr  
 la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos INSIDE\_narr  
 porte N Nc Nfs porte porte porte porte porte Ncfs Min NOM pos INSIDE\_narr

```

. F . . . . . F Min SENT punctuation INSIDE_narr
\n - - - - - Min - punctuation INSIDE_narr
toc N Np Nms toc toc toc toc Npms Min INT pos INSIDE_narr
, F , , , , , F Min PUN punctuation INSIDE_narr
toc N Np Nms toc toc toc toc toc Npms Min INT pos INSIDE_narr
. F . . . . . F Min SENT punctuation INSIDE_narr
\n\n - - - - - Min - punctuation INSIDE_narr
- F - - - - - F Min PUN punctuation BEGIN_spkr
" F " " " " " F Min PUN: cit punctuation INSIDE_spkr
qui P Pr P-ms qui P- qui P- qui Pr-ms- Min PRO:REL pos INSIDE_spkr
est V Vm V3-s est est Vip est est Vmip3s- Min VER:pres pos INSIDE_spkr
là R Rg Rp là là là là Rgp Min ADV pos INSIDE_spkr
? F ? ? ? ? ? F Min SENT punctuation INSIDE_spkr
" F " " " " " F Min PUN: cit punctuation INSIDE_spkr
. F . . . . . F Min SENT punctuation INSIDE_spkr
\n - - - - - Min - punctuation INSIDE_spkr
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos BEGIN_narr
petit A Af Ams Ap petit petit petit petit Afpms Min ADJ pos INSIDE_narr
chaperon N Nc Nms chaperon chaperon chaperon chaperon Ncms Min NOM pos INSIDE_narr
rouge A Af Ams Ap rouge rouge rouge rouge Afpms Min ADJ pos INSIDE_narr
, F , , , , , F Min PUN punctuation INSIDE_narr
qui P Pr P-ms qui P- qui P- qui Pr-ms- Min PRO:REL pos INSIDE_narr
entendit V Vm V3-s entendit entendit Vis entendit entendit Vmis3s- Min VER:simp pos INSIDE_narr
la D Da D-fs la D- la la Dd Da-fs-d Min DET:ART pos INSIDE_narr
grosse A Af Afs Ap grosse grosse grosse grosse Afpfs Min ADJ pos INSIDE_narr
voix N Nc Nfs voix voix voix voix Ncfs Min NOM pos INSIDE_narr
du S+D Sp+Da du+D-ms du+du du+D- du+du du+du du+Dd Sp+Da-ms-d Min PRP:det pos INSIDE_narr
loup N Nc Nms loup loup loup loup Ncms Min NOM pos INSIDE_narr
, F , , , , , F Min PUN punctuation INSIDE_narr
eut V Vm V3-s eut eut Vis eut eut Vmis3s- Min VER:simp pos INSIDE_narr
peur N Nc Nfs peur peur peur peur Ncfs Min NOM pos INSIDE_narr
d'abord A Af Afs Ap d'abord d'abord d'abord d'abord Afpfs Min ADV pos INSIDE_narr
, F , , , , , F Min PUN punctuation INSIDE_narr
mais C Cc mais mais mais mais mais Cc Min KON pos INSIDE_narr
croyant V Vm V- croyant croyant Vpp croyant croyant Vmpp- Min VER:ppre pos INSIDE_narr
que C Cs que que que que que Cs Min KON pos INSIDE_narr
sa D Ds D3fs sa Ds sa sa D- Ds3fss- Min DET:POS pos INSIDE_narr
mère-grand N Nc Nfs mère-grand mère-grand mère-grand mère-grand Ncfs Min NOM pos INSIDE_narr
était V Va V3-s était était Vii était était Vai3s- Min VER:impf pos INSIDE_narr
enrhumée V Vm V-ms enrhumée enrhumée Vps enrhumée enrhumée Vmps-sm Min VER:pper pos INSIDE_narr
, F , , , , , F Min PUN punctuation INSIDE_narr
répondit V Vm V3-s répondit répondit Vis répondit répondit Vmis3s- Min VER:simp pos INSIDE_narr
: F : : : : : F Min PUN punctuation INSIDE_narr
\n - - - - - Min - punctuation INSIDE_narr
- F - - - - - F Min PUN punctuation BEGIN_spkr
" F " " " " " F Min PUN: cit punctuation INSIDE_spkr
c' P Pd P-ms c' P- c' P- c' Pd-ms- Min PRO:DEM pos INSIDE_spkr
est V Vm V3-s est est Vip est est Vmip3s- Min VER:pres pos INSIDE_spkr
votre D Ds D2fs votre Dp votre D- Ds2fsp- Min DET:POS pos INSIDE_spkr
fille N Nc Nfs fille fille fille fille fille Ncfs Min NOM pos INSIDE_spkr
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos INSIDE_spkr
petit A Af Ams Ap petit petit petit petit Afpms Min ADJ pos INSIDE_spkr
chaperon N Nc Nms chaperon chaperon chaperon chaperon Ncms Min NOM pos INSIDE_spkr
rouge A Af Ams Ap rouge rouge rouge rouge Afpms Min ADJ pos INSIDE_spkr
, F , , , , , F Min PUN punctuation INSIDE_spkr
qui P Pr P-ms qui P- qui P- qui Pr-ms- Min PRO:REL pos INSIDE_spkr
vous P Pp P2mp vous P- vous Pa vous Pp2mpa- Min PRO:PER pos INSIDE_spkr
apporte V Vm V3-s apporte apporte Vip apporte apporte Vmip3s- Min VER:pres pos INSIDE_spkr
une D Da D-fs une D- une une Di Da-fs-i Min DET:ART pos INSIDE_spkr
galette N Nc Nfs galette galette galette galette Ncfs Min NOM pos INSIDE_spkr
et C Cc et et et et et et Cc Min KON pos INSIDE_spkr
un D Da D-ms un D- un un Di Da-ms-i Min DET:ART pos INSIDE_spkr
petit A Af Ams Ap petit petit petit petit Afpms Min ADJ pos INSIDE_spkr
pot N Nc Nms pot pot pot pot Ncms Min NOM pos INSIDE_spkr
de S Sp de de de de de de Sp Min PRP pos INSIDE_spkr
beurre N Nc Nms beurre beurre beurre beurre Ncms Min NOM pos INSIDE_spkr
que C Cs que que que que Cs Min KON pos INSIDE_spkr
ma D Ds D1fs ma Ds ma ma D- Ds1fss- Min DET:POS pos INSIDE_spkr
mère N Nc Nfs mère mère mère mère Ncfs Min NOM pos INSIDE_spkr
vous P Pp P2mp vous P- vous Pa vous Pp2mpa- Min PRO:PER pos INSIDE_spkr
envoie V Vm V1-s envoie envoie Vip envoie envoie Vmip1s- Min VER:pres pos INSIDE_spkr
. F . . . . . F Min SENT punctuation INSIDE_spkr
" F " " " " " F Min PUN: cit punctuation INSIDE_spkr
\n - - - - - Min - punctuation INSIDE_spkr
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos BEGIN_narr
loup N Nc Nms loup loup loup loup loup Ncms Min NOM pos INSIDE_narr
lui P Pp P3ms lui P- lui Pd lui Pp3msd- Min PRO:PER pos INSIDE_narr
cria V Vm V3-s cria cria Vis cria cria Vmis3s- Min VER:simp pos INSIDE_narr

```

, F , , , , , , F Min PUN punctuation INSIDE\_narr  
en S Sp en en en en en en Sp Min PRP pos INSIDE\_narr  
adoucissant V Vm V— adoucissant adoucissant Vn— adoucissant adoucissant Vmn— Min VER:ppre pos INSIDE\_narr  
un D Da D—ms un D— un un Di Da—ms—i Min DET:ART pos INSIDE\_narr  
peu R Rg Rp peu peu peu peu peu Rgp Min ADV pos INSIDE\_narr  
sa D Ds D3fs sa Ds sa sa D— Ds3fs— Min DET:POS pos INSIDE\_narr  
voix N Nc Nfs voix voix voix voix voix Ncfs Min NOM pos INSIDE\_narr  
: F : : : : : : F Min PUN punctuation INSIDE\_narr  
\n — — — — — — — Min — punctuation INSIDE\_narr  
— F — — — — — — F Min PUN punctuation BEGIN\_spkr  
" F " " " " " " F Min PUN:cit punctuation INSIDE\_spkr  
tire V Vm V3—s tire tire Vip tire tire Vmip3s— Min VER:pres pos INSIDE\_spkr  
la D Da D—fs la D— la la Dd Da—fs—d Min DET:ART pos INSIDE\_spkr  
chevillette N Nc Nfs chevillette chevillette chevillette chevillette chevillette Ncfs Min NOM pos INSIDE\_spkr  
, F , , , , , , F Min PUN punctuation INSIDE\_spkr  
la D Da D—fs la D— la la Dd Da—fs—d Min DET:ART pos INSIDE\_spkr  
bobinette N Nc Nfs bobinette bobinette bobinette bobinette bobinette Ncfs Min NOM pos INSIDE\_spkr  
cherra A Af Afs Ap cherra cherra cherra cherra Afpfs Min VER:simp pos INSIDE\_spkr  
. F . . . . . F Min SENT punctuation INSIDE\_spkr  
" F " " " " " " F Min PUN:cit punctuation INSIDE\_spkr  
\n \n — — — — — — — Min — punctuation INSIDE\_spkr  
le D Da D—ms le D— le le Dd Da—ms—d Min DET:ART pos BEGIN\_narr  
petit A Af Ams Ap petit petit petit petit Afpms Min ADJ pos INSIDE\_narr  
chaperon N Nc Nms chaperon chaperon chaperon chaperon chaperon Ncms Min NOM pos INSIDE\_narr  
rouge A Af Ams Ap rouge rouge rouge rouge Afpms Min ADJ pos INSIDE\_narr  
tira V Vm V3—s tira tira Vis tira tira Vmis3s— Min VER:simp pos INSIDE\_narr  
la D Da D—fs la D— la la Dd Da—fs—d Min DET:ART pos INSIDE\_narr  
chevillette N Nc Nfs chevillette chevillette chevillette chevillette chevillette Ncfs Min NOM pos INSIDE\_narr  
, F , , , , , , F Min PUN punctuation INSIDE\_narr  
et C Cc et et et et et et Cc Min KON pos INSIDE\_narr  
la D Da D—fs la D— la la Dd Da—fs—d Min DET:ART pos INSIDE\_narr  
porte N Nc Nfs porte porte porte porte Ncfs Min NOM pos INSIDE\_narr  
s' P Px P3fs s' P— s' P— s' Px3fs— Min PRO:PER pos INSIDE\_narr  
ouvrit V Vm V3—s ouvrit ouvrit Vis ouvrit ouvrit Vmis3s— Min VER:simp pos INSIDE\_narr  
. F . . . . . F Min SENT punctuation INSIDE\_narr  
\n — — — — — — — Min — punctuation INSIDE\_narr  
le D Da D—ms le D— le le Dd Da—ms—d Min DET:ART pos INSIDE\_narr  
loup N Nc Nms loup loup loup loup loup Ncms Min NOM pos INSIDE\_narr  
, F , , , , , , F Min PUN punctuation INSIDE\_narr  
la P Pp P3fs la P— la Pa la Pp3fsa— Min PRO:PER pos INSIDE\_narr  
voyant V Vm V— voyant voyant Vpp voyant voyant Vmpp— Min VER:ppre pos INSIDE\_narr  
entrer V Vm V— entrer entrer Vn— entrer entrer Vmn— Min VER:infi pos INSIDE\_narr  
, F , , , , , , F Min PUN punctuation INSIDE\_narr  
lui P Pp P3ms lui P— lui Pd lui Pp3msd— Min PRO:PER pos INSIDE\_narr  
dit V Vm V3—s dit dit Vip dit dit Vmip3s— Min VER:pres pos INSIDE\_narr  
en S Sp en en en en en en Sp Min PRP pos INSIDE\_narr  
se P Px P3ms se P— se P— se Px3ms— Min PRO:PER pos INSIDE\_narr  
cachant V Vm V— cachant cachant Vpp cachant cachant Vmpp— Min VER:ppre pos INSIDE\_narr  
dans S Sp dans dans dans dans dans Sp Min PRP pos INSIDE\_narr  
le D Da D—ms le D— le le Dd Da—ms—d Min DET:ART pos INSIDE\_narr  
lit N Nc Nms lit lit lit lit lit Ncms Min NOM pos INSIDE\_narr  
sous S Sp sous sous sous sous sous Sp Min PRP pos INSIDE\_narr  
la D Da D—fs la D— la la Dd Da—fs—d Min DET:ART pos INSIDE\_narr  
couverture N Nc Nfs couverture couverture couverture couverture couverture Ncfs Min NOM pos INSIDE\_narr  
: F : : : : : : F Min PUN punctuation INSIDE\_narr  
\n — — — — — — — Min — punctuation INSIDE\_narr  
— F — — — — — — F Min PUN punctuation BEGIN\_spkr  
" F " " " " " " F Min PUN:cit punctuation INSIDE\_spkr  
mets V Vm V2—s mets mets Vmp mets mets Vmmp2s— Min VER:pres pos INSIDE\_spkr  
la D Da D—fs la D— la la Dd Da—fs—d Min DET:ART pos INSIDE\_spkr  
galette N Nc Nfs galette galette galette galette galette Ncfs Min NOM pos INSIDE\_spkr  
et C Cc et et et et et et Cc Min KON pos INSIDE\_spkr  
le D Da D—ms le D— le le Dd Da—ms—d Min DET:ART pos INSIDE\_spkr  
petit A Af Ams Ap petit petit petit petit Afpms Min ADJ pos INSIDE\_spkr  
pot N Nc Nms pot pot pot pot pot Ncms Min NOM pos INSIDE\_spkr  
de S Sp de de de de de de Sp Min PRP pos INSIDE\_spkr  
beurre N Nc Nms beurre beurre beurre beurre Ncms Min NOM pos INSIDE\_spkr  
sur S Sp sur sur sur sur sur Sp Min PRP pos INSIDE\_spkr  
la D Da D—fs la D— la la Dd Da—fs—d Min DET:ART pos INSIDE\_spkr  
huche N Nc Nfs huche huche huche huche huche Ncfs Min NOM pos INSIDE\_spkr  
, F , , , , , , F Min PUN punctuation INSIDE\_spkr  
et C Cc et et et et et et Cc Min KON pos INSIDE\_spkr  
viens V Vm V2—s viens viens Vmp viens viens Vmmp2s— Min VER:pres pos INSIDE\_spkr  
te P Pp P2ms te P— te Pd te Pp2msd— Min PRO:PER pos INSIDE\_spkr  
coucher V Vm V— coucher coucher Vn— coucher coucher Vmn— Min VER:infi pos INSIDE\_spkr  
avec S Sp avec avec avec avec avec avec Sp Min PRP pos INSIDE\_spkr  
moi P Pp Plms moi P— moi Po moi Pplms— Min PRO:PER pos INSIDE\_spkr  
. F . . . . . F Min SENT punctuation INSIDE\_spkr

## C.1. Données d'entrée

```

" F " " " " " " F Min PUN: cit punctuation INSIDE_spkr
\n\n - - - - - Min - punctuation INSIDE_spkr
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos BEGIN_narr
petit A Af Ams Ap petit petit petit petit Afpms Min ADJ pos INSIDE_narr
chaperon N Nc Nms chaperon chaperon chaperon chaperon chaperon Ncms Min NOM pos INSIDE_narr
rouge A Af Ams Ap rouge rouge rouge rouge Afpms Min ADJ pos INSIDE_narr
se P Px P3ms se P- se P- se Px3ms- Min PRO:PER pos INSIDE_narr
deshabille V Vm V- deshabille deshabille Vn- deshabille deshabille Vmn- Min VER:pres pos INSIDE_narr
, F , , , , , F Min PUN punctuation INSIDE_narr
et C Cc et et et et et Cc Min KON pos INSIDE_narr
va V Vm V3-s va va Vip va va Vmip3s- Min VER:pres pos INSIDE_narr
se P Px P3ms se P- se P- se Px3ms- Min PRO:PER pos INSIDE_narr
mettre V Vm V- mettre mettre Vn- mettre mettre Vmn- Min VER:infi pos INSIDE_narr
dans S Sp dans dans dans dans dans Sp Min PRP pos INSIDE_narr
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos INSIDE_narr
lit N Nc Nms lit lit lit lit lit Ncms Min NOM pos INSIDE_narr
, F , , , , , F Min PUN punctuation INSIDE_narr
où P Pr P-ms où P- où P- où Pr-ms- Min PRO:REL pos INSIDE_narr
elle P Pp P3fs elle P- elle Pn elle Pp3fsn- Min PRO:PER pos INSIDE_narr
fut V Va V3-s fut fut Vis fut fut Vais3s- Min VER:simp pos INSIDE_narr
bien R Rg Rp bien bien bien bien Rgp Min ADV pos INSIDE_narr
étonnée V Vm V-fs étonnée étonnée Vps étonnée étonnée Vmps-sf Min VER:pper pos INSIDE_narr
de S Sp de de de de de Sp Min PRP pos INSIDE_narr
voir V Vm V- voir voir Vn- voir voir Vmn- Min VER:infi pos INSIDE_narr
comment R Rx Rp comment comment comment comment comment Rxp Min ADV pos INSIDE_narr
sa D Ds D3fs sa Ds sa sa D- Ds3fs- Min DET:POS pos INSIDE_narr
mère-grand N Nc Nfs mère-grand mère-grand mère-grand mère-grand mère-grand Ncfs Min NOM pos INSIDE_narr
était V Va V3-s était était Vii était était Viii3s- Min VER:impf pos INSIDE_narr
faite V Vm V-fs faite faite Vps faite faite Vmps-sf Min VER:pper pos INSIDE_narr
en S Sp en en en en en Sp Min PRP pos INSIDE_narr
son D Ds D3ms son Ds son son D- Ds3ms- Min DET:POS pos INSIDE_narr
deshabillé A Af Ams Ap deshabillé deshabillé deshabillé deshabillé Afpms Min NOM pos INSIDE_narr
. F . . . . . F Min SENT punctuation INSIDE_narr
\n - - - - - Min - punctuation INSIDE_narr
elle P Pp P3fs elle P- elle Pn elle Pp3fsn- Min PRO:PER pos INSIDE_narr
lui P Pp P3ms lui P- lui Pd lui Pp3msd- Min PRO:PER pos INSIDE_narr
dit V Vm V3-s dit dit Vip dit dit Vmip3s- Min VER:pres pos INSIDE_narr
: F : : : : : F Min PUN punctuation INSIDE_narr
\n - - - - - Min - punctuation INSIDE_narr
- F - - - - - F Min PUN punctuation BEGIN_spkr
" F " " " " " " F Min PUN: cit punctuation INSIDE_spkr
ma D Ds D1fs ma Ds ma ma D- Ds1fs- Min DET:POS pos INSIDE_spkr
mère-grand N Nc Nfs mère-grand mère-grand mère-grand mère-grand mère-grand Ncfs Min NOM pos INSIDE_spkr
que C Cs que que que que que Cs Min KON pos INSIDE_spkr
vous P Pp P2mp vous P- vous Pn vous Pp2mpn- Min PRO:PER pos INSIDE_spkr
avez V Vm V2-p avez avez Vip avez avez Vmip2p- Min VER:pres pos INSIDE_spkr
de S Sp de de de de de de Sp Min PRP pos INSIDE_spkr
grands A Af Amp Ap grands grands grands grands Afpmp Min ADJ pos INSIDE_spkr
bras N Nc Nmp bras bras bras bras Ncmp Min NOM pos INSIDE_spkr
! F ! ! ! ! ! F Min SENT punctuation INSIDE_spkr
" F " " " " " " F Min PUN: cit punctuation INSIDE_spkr
\n - - - - - Min - punctuation INSIDE_spkr
- F - - - - - F Min PUN punctuation BEGIN_spkr
" F " " " " " " F Min PUN: cit punctuation INSIDE_spkr
c' P Pd P-ms c' P- c' P- c' Pd-ms- Min PRO:DEM pos INSIDE_spkr
est V Vm V3-s est est Vip est est Vmip3s- Min VER:pres pos INSIDE_spkr
pour S Sp pour pour pour pour pour Sp Min KON pos INSIDE_spkr
mieux R Rg Rc mieux mieux mieux mieux Rgc Min ADV pos INSIDE_spkr
t' P Pp P2ms t' P- t' Pa t' Pp2msa- Min PRO:PER pos INSIDE_spkr
embrasser V Vm V- embrasser embrasser Vn- embrasser embrasser Vmn- Min VER:infi pos INSIDE_spkr
ma D Ds D1fs ma Ds ma ma D- Ds1fs- Min DET:POS pos INSIDE_spkr
fille N Nc Nfs fille fille fille fille fille Ncfs Min NOM pos INSIDE_spkr
. F . . . . . F Min SENT punctuation INSIDE_spkr
" F " " " " " " F Min PUN: cit punctuation INSIDE_spkr
\n - - - - - Min - punctuation INSIDE_spkr
- F - - - - - F Min PUN punctuation BEGIN_spkr
" F " " " " " " F Min PUN: cit punctuation INSIDE_spkr
ma D Ds D1fs ma Ds ma ma D- Ds1fs- Min DET:POS pos INSIDE_spkr
mère-grand N Nc Nfs mère-grand mère-grand mère-grand mère-grand mère-grand Ncfs Min NOM pos INSIDE_spkr
que C Cs que que que que que Cs Min KON pos INSIDE_spkr
vous P Pp P2mp vous P- vous Pn vous Pp2mpn- Min PRO:PER pos INSIDE_spkr
avez V Vm V2-p avez avez Vip avez avez Vmip2p- Min VER:pres pos INSIDE_spkr
de S Sp de de de de de de Sp Min PRP pos INSIDE_spkr
grandes A Af Afp Ap grandes grandes grandes grandes Afpfp Min ADJ pos INSIDE_spkr
jambes N Nc Nfp jambes jambes jambes jambes jambes Ncfp Min NOM pos INSIDE_spkr
! F ! ! ! ! ! F Min SENT punctuation INSIDE_spkr
" F " " " " " " F Min PUN: cit punctuation INSIDE_spkr
\n - - - - - Min - punctuation INSIDE_spkr

```



```

- F - - - - - F Min PUN punctuation BEGIN_spkr
" F " " " " " " F Min PUN:cit punctuation INSIDE_spkr
c' P Pd P-ms c' P- c' P- c' Pd-ms- Min PRO:DEM pos INSIDE_spkr
est V Vm V3-s est est Vip est est Vmip3s- Min VER:pres pos INSIDE_spkr
pour S Sp pour pour pour pour pour pour Sp Min KON pos INSIDE_spkr
mieux R Rg Rc mieux mieux mieux mieux mieux Rgc Min ADV pos INSIDE_spkr
courir V Vm V- courir courir Vn- courir courir Vmn- Min VER:infi pos INSIDE_spkr
mon D Ds Dlms mon Ds mon mon D- Dslmss- Min DET:POS pos INSIDE_spkr
enfant N Nc Nms enfant enfant enfant enfant enfant Ncms Min NOM pos INSIDE_spkr
. F . . . . . F Min SENT punctuation INSIDE_spkr
" F " " " " " " F Min PUN:cit punctuation INSIDE_spkr
\n - - - - - Min - punctuation INSIDE_spkr
- F - - - - - F Min PUN punctuation BEGIN_spkr
" F " " " " " " F Min PUN:cit punctuation INSIDE_spkr
ma D Ds D1fs ma Ds ma ma D- Ds1fss- Min DET:POS pos INSIDE_spkr
mère-grand N Nc Nfs mère-grand mère-grand mère-grand mère-grand Ncfs Min NOM pos INSIDE_spkr
que C Cs que que que que que Cs Min KON pos INSIDE_spkr
vous P Pp P2mp vous P- vous Pn vous Pp2mpn- Min PRO:PER pos INSIDE_spkr
avez V Vm V2-p avez avez Vip avez avez Vmip2p- Min VER:pres pos INSIDE_spkr
de S Sp de de de de de de Sp Min PRP pos INSIDE_spkr
grandes A Af Afp Ap grandes grandes grandes grandes Afpfp Min ADJ pos INSIDE_spkr
oreilles N Nc Nfp oreilles oreilles oreilles oreilles oreilles Ncfp Min NOM pos INSIDE_spkr
! F ! ! ! ! ! ! F Min SENT punctuation INSIDE_spkr
" F " " " " " " F Min PUN:cit punctuation INSIDE_spkr
\n - - - - - Min - punctuation INSIDE_spkr
- F - - - - - F Min PUN punctuation BEGIN_spkr
" F " " " " " " F Min PUN:cit punctuation INSIDE_spkr
c' P Pd P-ms c' P- c' P- c' Pd-ms- Min PRO:DEM pos INSIDE_spkr
est V Vm V3-s est est Vip est est Vmip3s- Min VER:pres pos INSIDE_spkr
pour S Sp pour pour pour pour pour pour Sp Min KON pos INSIDE_spkr
mieux R Rg Rc mieux mieux mieux mieux mieux Rgc Min ADV pos INSIDE_spkr
écouter V Vm V- écouter écouter Vn- écouter écouter Vmn- Min VER:infi pos INSIDE_spkr
mon D Ds Dlms mon Ds mon mon D- Dslmss- Min DET:POS pos INSIDE_spkr
enfant N Nc Nms enfant enfant enfant enfant enfant Ncms Min NOM pos INSIDE_spkr
. F . . . . . F Min SENT punctuation INSIDE_spkr
" F " " " " " " F Min PUN:cit punctuation INSIDE_spkr
\n - - - - - Min - punctuation INSIDE_spkr
- F - - - - - F Min PUN punctuation BEGIN_spkr
" F " " " " " " F Min PUN:cit punctuation INSIDE_spkr
ma D Ds D1fs ma Ds ma ma D- Ds1fss- Min DET:POS pos INSIDE_spkr
mère-grand N Nc Nfs mère-grand mère-grand mère-grand mère-grand Ncfs Min NOM pos INSIDE_spkr
que C Cs que que que que que Cs Min KON pos INSIDE_spkr
vous P Pp P2mp vous P- vous Pn vous Pp2mpn- Min PRO:PER pos INSIDE_spkr
avez V Vm V2-p avez avez Vip avez avez Vmip2p- Min VER:pres pos INSIDE_spkr
de S Sp de de de de de de Sp Min PRP pos INSIDE_spkr
grands A Af Amp Ap grands grands grands grands Afpmp Min ADJ pos INSIDE_spkr
yeux N Nc Nmp yeux yeux yeux yeux yeux Ncmp Min NOM pos INSIDE_spkr
! F ! ! ! ! ! ! F Min SENT punctuation INSIDE_spkr
" F " " " " " " F Min PUN:cit punctuation INSIDE_spkr
\n - - - - - Min - punctuation INSIDE_spkr
- F - - - - - F Min PUN punctuation BEGIN_spkr
" F " " " " " " F Min PUN:cit punctuation INSIDE_spkr
c' P Pd P-ms c' P- c' P- c' Pd-ms- Min PRO:DEM pos INSIDE_spkr
est V Vm V3-s est est Vip est est Vmip3s- Min VER:pres pos INSIDE_spkr
pour S Sp pour pour pour pour pour pour Sp Min KON pos INSIDE_spkr
mieux R Rg Rc mieux mieux mieux mieux mieux Rgc Min ADV pos INSIDE_spkr
voir V Vm V- voir voir Vn- voir voir Vmn- Min VER:infi pos INSIDE_spkr
mon D Ds Dlms mon Ds mon mon D- Dslmss- Min DET:POS pos INSIDE_spkr
enfant N Nc Nms enfant enfant enfant enfant enfant Ncms Min NOM pos INSIDE_spkr
. F . . . . . F Min SENT punctuation INSIDE_spkr
" F " " " " " " F Min PUN:cit punctuation INSIDE_spkr
\n - - - - - Min - punctuation INSIDE_spkr
- F - - - - - F Min PUN punctuation BEGIN_spkr
" F " " " " " " F Min PUN:cit punctuation INSIDE_spkr
ma D Ds D1fs ma Ds ma ma D- Ds1fss- Min DET:POS pos INSIDE_spkr
mère-grand N Nc Nfs mère-grand mère-grand mère-grand mère-grand Ncfs Min NOM pos INSIDE_spkr
que C Cs que que que que que Cs Min KON pos INSIDE_spkr
vous P Pp P2mp vous P- vous Pn vous Pp2mpn- Min PRO:PER pos INSIDE_spkr
avez V Vm V2-p avez avez Vip avez avez Vmip2p- Min VER:pres pos INSIDE_spkr
de S Sp de de de de de de Sp Min PRP pos INSIDE_spkr
grandes A Af Afp Ap grandes grandes grandes grandes Afpfp Min ADJ pos INSIDE_spkr
dents N Nc Nfp dents dents dents dents dents Ncfp Min NOM pos INSIDE_spkr
! F ! ! ! ! ! ! F Min SENT punctuation INSIDE_spkr
" F " " " " " " F Min PUN:cit punctuation INSIDE_spkr
\n - - - - - Min - punctuation INSIDE_spkr
- F - - - - - F Min PUN punctuation BEGIN_spkr
" F " " " " " " F Min PUN:cit punctuation INSIDE_spkr
c' P Pd P-ms c' P- c' P- c' Pd-ms- Min PRO:DEM pos INSIDE_spkr

```

```

est V Vm V3-s est est Vip est est Vmip3s- Min VER:pres pos      INSIDE_spkr
pour S Sp pour pour pour pour pour Sp Min PRP pos      INSIDE_spkr
te P Pp P2ms te P- te Pd te Pp2msd- Min PRO:PER pos      INSIDE_spkr
manger V Vm V- manger manger Vn- manger manger Vmn- Min VER:infi pos      INSIDE_spkr
. F . . . . . F Min SENT punctuation      INSIDE_spkr
" F " " " " " " F Min PUN:cit punctuation      INSIDE_spkr
\n\n - - - - - Min - punctuation      INSIDE_spkr
et C Cc et et et et et et Cc Min KON pos      BEGIN_narr
en S Sp en en en en en en Sp Min PRO:PER pos      INSIDE_narr
disant V Vm V- disant disant Vpp disant disant Vmpp- Min VER:ppe pos      INSIDE_narr
ces D Dd D-mp ces D- ces ces D- Dd-mp- Min PRO:DEM pos      INSIDE_narr
mots N Nc Nmp mots mots mots mots mots Ncmp Min NOM pos      INSIDE_narr
, F , , , , , F Min PUN punctuation      INSIDE_narr
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos      INSIDE_narr
méchant A Af Ams Ap méchant méchant méchant méchant Afms Min ADJ pos      INSIDE_narr
loup N Nc Nms loup loup loup loup loup Ncms Min NOM pos      INSIDE_narr
se P Px P3ms se P- se P- se Px3ms- Min PRO:PER pos      INSIDE_narr
jeta V Vm V3-s jeta jeta Vis jeta jeta Vmis3s- Min VER:simp pos      INSIDE_narr
sur S Sp sur sur sur sur sur Sp Min PRP pos      INSIDE_narr
le D Da D-ms le D- le le Dd Da-ms-d Min DET:ART pos      INSIDE_narr
petit A Af Ams Ap petit petit petit petit Afms Min ADJ pos      INSIDE_narr
chaperon N Nc Nms chaperon chaperon chaperon chaperon Ncms Min NOM pos      INSIDE_narr
rouge A Af Ams Ap rouge rouge rouge rouge Afms Min ADJ pos      INSIDE_narr
, F , , , , , F Min PUN punctuation      INSIDE_narr
et C Cc et et et et et et Cc Min KON pos      INSIDE_narr
la P Pp P3fs la P- la Pa la Pp3fsa- Min PRO:PER pos      INSIDE_narr
mangea V Vm V3-s mangea mangea Vis mangea mangea Vmis3s- Min VER:simp pos      INSIDE_narr
. F . . . . . F Min SENT punctuation      INSIDE_narr
\n - - - - - Min - punctuation      INSIDE_narr

```

## C.2 Patron d'extraction de fonction features

Le patron détaillé ci-dessous consiste à considérer chaque unigramme et chaque bigramme d'annotation de tours de parole. Les 13 dimensions des observations sont prises en compte dans une fenêtre -9,+9 autour de l'observation courante. Les observations sont également regroupées sous forme de bigramme dans une fenêtre -2,+2 autour de l'observation courante.

```

*1gram1:%x[-9,0]
*1gram2:%x[-9,1]
*1gram3:%x[-9,2]
*1gram4:%x[-9,3]
*1gram5:%x[-9,4]
*1gram6:%x[-9,5]
*1gram7:%x[-9,6]
*1gram8:%x[-9,7]
*1gram9:%x[-9,8]
*1gram10:%x[-9,9]
*1gram11:%x[-9,10]
*1gram12:%x[-9,11]
*1gram13:%x[-9,12]
*1gram14:%x[-8,0]
*1gram15:%x[-8,1]
*1gram16:%x[-8,2]
*1gram17:%x[-8,3]
*1gram18:%x[-8,4]
*1gram19:%x[-8,5]
*1gram20:%x[-8,6]
*1gram21:%x[-8,7]
*1gram22:%x[-8,8]
*1gram23:%x[-8,9]
*1gram24:%x[-8,10]
*1gram25:%x[-8,11]
*1gram26:%x[-8,12]
*1gram27:%x[-7,0]
*1gram28:%x[-7,1]
*1gram29:%x[-7,2]
*1gram30:%x[-7,3]
*1gram31:%x[-7,4]
*1gram32:%x[-7,5]
*1gram33:%x[-7,6]

```

```
*lgram34:%x[-7,7]
*lgram35:%x[-7,8]
*lgram36:%x[-7,9]
*lgram37:%x[-7,10]
*lgram38:%x[-7,11]
*lgram39:%x[-7,12]
*lgram40:%x[-6,0]
*lgram41:%x[-6,1]
*lgram42:%x[-6,2]
*lgram43:%x[-6,3]
*lgram44:%x[-6,4]
*lgram45:%x[-6,5]
*lgram46:%x[-6,6]
*lgram47:%x[-6,7]
*lgram48:%x[-6,8]
*lgram49:%x[-6,9]
*lgram50:%x[-6,10]
*lgram51:%x[-6,11]
*lgram52:%x[-6,12]
*lgram53:%x[-5,0]
*lgram54:%x[-5,1]
*lgram55:%x[-5,2]
*lgram56:%x[-5,3]
*lgram57:%x[-5,4]
*lgram58:%x[-5,5]
*lgram59:%x[-5,6]
*lgram60:%x[-5,7]
*lgram61:%x[-5,8]
*lgram62:%x[-5,9]
*lgram63:%x[-5,10]
*lgram64:%x[-5,11]
*lgram65:%x[-5,12]
*lgram66:%x[-4,0]
*lgram67:%x[-4,1]
*lgram68:%x[-4,2]
*lgram69:%x[-4,3]
*lgram70:%x[-4,4]
*lgram71:%x[-4,5]
*lgram72:%x[-4,6]
*lgram73:%x[-4,7]
*lgram74:%x[-4,8]
*lgram75:%x[-4,9]
*lgram76:%x[-4,10]
*lgram77:%x[-4,11]
*lgram78:%x[-4,12]
*lgram79:%x[-3,0]
*lgram80:%x[-3,1]
*lgram81:%x[-3,2]
*lgram82:%x[-3,3]
*lgram83:%x[-3,4]
*lgram84:%x[-3,5]
*lgram85:%x[-3,6]
*lgram86:%x[-3,7]
*lgram87:%x[-3,8]
*lgram88:%x[-3,9]
*lgram89:%x[-3,10]
*lgram90:%x[-3,11]
*lgram91:%x[-3,12]
*lgram92:%x[-2,0]
*lgram93:%x[-2,1]
*lgram94:%x[-2,2]
*lgram95:%x[-2,3]
*lgram96:%x[-2,4]
*lgram97:%x[-2,5]
*lgram98:%x[-2,6]
*lgram99:%x[-2,7]
*lgram100:%x[-2,8]
*lgram101:%x[-2,9]
*lgram102:%x[-2,10]
*lgram103:%x[-2,11]
*lgram104:%x[-2,12]
*lgram105:%x[-1,0]
*lgram106:%x[-1,1]
*lgram107:%x[-1,2]
*lgram108:%x[-1,3]
*lgram109:%x[-1,4]
*lgram110:%x[-1,5]
*lgram111:%x[-1,6]
```

```
*lgram112:%x[-1,7]
*lgram113:%x[-1,8]
*lgram114:%x[-1,9]
*lgram115:%x[-1,10]
*lgram116:%x[-1,11]
*lgram117:%x[-1,12]
*lgram118:%x[0,0]
*lgram119:%x[0,1]
*lgram120:%x[0,2]
*lgram121:%x[0,3]
*lgram122:%x[0,4]
*lgram123:%x[0,5]
*lgram124:%x[0,6]
*lgram125:%x[0,7]
*lgram126:%x[0,8]
*lgram127:%x[0,9]
*lgram128:%x[0,10]
*lgram129:%x[0,11]
*lgram130:%x[0,12]
*lgram131:%x[1,0]
*lgram132:%x[1,1]
*lgram133:%x[1,2]
*lgram134:%x[1,3]
*lgram135:%x[1,4]
*lgram136:%x[1,5]
*lgram137:%x[1,6]
*lgram138:%x[1,7]
*lgram139:%x[1,8]
*lgram140:%x[1,9]
*lgram141:%x[1,10]
*lgram142:%x[1,11]
*lgram143:%x[1,12]
*lgram144:%x[2,0]
*lgram145:%x[2,1]
*lgram146:%x[2,2]
*lgram147:%x[2,3]
*lgram148:%x[2,4]
*lgram149:%x[2,5]
*lgram150:%x[2,6]
*lgram151:%x[2,7]
*lgram152:%x[2,8]
*lgram153:%x[2,9]
*lgram154:%x[2,10]
*lgram155:%x[2,11]
*lgram156:%x[2,12]
*lgram157:%x[3,0]
*lgram158:%x[3,1]
*lgram159:%x[3,2]
*lgram160:%x[3,3]
*lgram161:%x[3,4]
*lgram162:%x[3,5]
*lgram163:%x[3,6]
*lgram164:%x[3,7]
*lgram165:%x[3,8]
*lgram166:%x[3,9]
*lgram167:%x[3,10]
*lgram168:%x[3,11]
*lgram169:%x[3,12]
*lgram170:%x[4,0]
*lgram171:%x[4,1]
*lgram172:%x[4,2]
*lgram173:%x[4,3]
*lgram174:%x[4,4]
*lgram175:%x[4,5]
*lgram176:%x[4,6]
*lgram177:%x[4,7]
*lgram178:%x[4,8]
*lgram179:%x[4,9]
*lgram180:%x[4,10]
*lgram181:%x[4,11]
*lgram182:%x[4,12]
*lgram183:%x[5,0]
*lgram184:%x[5,1]
*lgram185:%x[5,2]
*lgram186:%x[5,3]
*lgram187:%x[5,4]
*lgram188:%x[5,5]
*lgram189:%x[5,6]
```

```

*1gram190:%x[5,7]
*1gram191:%x[5,8]
*1gram192:%x[5,9]
*1gram193:%x[5,10]
*1gram194:%x[5,11]
*1gram195:%x[5,12]
*1gram196:%x[6,0]
*1gram197:%x[6,1]
*1gram198:%x[6,2]
*1gram199:%x[6,3]
*1gram200:%x[6,4]
*1gram201:%x[6,5]
*1gram202:%x[6,6]
*1gram203:%x[6,7]
*1gram204:%x[6,8]
*1gram205:%x[6,9]
*1gram206:%x[6,10]
*1gram207:%x[6,11]
*1gram208:%x[6,12]
*1gram209:%x[7,0]
*1gram210:%x[7,1]
*1gram211:%x[7,2]
*1gram212:%x[7,3]
*1gram213:%x[7,4]
*1gram214:%x[7,5]
*1gram215:%x[7,6]
*1gram216:%x[7,7]
*1gram217:%x[7,8]
*1gram218:%x[7,9]
*1gram219:%x[7,10]
*1gram220:%x[7,11]
*1gram221:%x[7,12]
*1gram222:%x[8,0]
*1gram223:%x[8,1]
*1gram224:%x[8,2]
*1gram225:%x[8,3]
*1gram226:%x[8,4]
*1gram227:%x[8,5]
*1gram228:%x[8,6]
*1gram229:%x[8,7]
*1gram230:%x[8,8]
*1gram231:%x[8,9]
*1gram232:%x[8,10]
*1gram233:%x[8,11]
*1gram234:%x[8,12]
*1gram235:%x[9,0]
*1gram236:%x[9,1]
*1gram237:%x[9,2]
*1gram238:%x[9,3]
*1gram239:%x[9,4]
*1gram240:%x[9,5]
*1gram241:%x[9,6]
*1gram242:%x[9,7]
*1gram243:%x[9,8]
*1gram244:%x[9,9]
*1gram245:%x[9,10]
*1gram246:%x[9,11]
*1gram247:%x[9,12]
*2gram248:%x[-2,0]/%x[-1,0]
*2gram249:%x[-2,1]/%x[-1,1]
*2gram250:%x[-2,2]/%x[-1,2]
*2gram251:%x[-2,3]/%x[-1,3]
*2gram252:%x[-2,4]/%x[-1,4]
*2gram253:%x[-2,5]/%x[-1,5]
*2gram254:%x[-2,6]/%x[-1,6]
*2gram255:%x[-2,7]/%x[-1,7]
*2gram256:%x[-2,8]/%x[-1,8]
*2gram257:%x[-2,9]/%x[-1,9]
*2gram258:%x[-2,10]/%x[-1,10]
*2gram259:%x[-2,11]/%x[-1,11]
*2gram260:%x[-2,12]/%x[-1,12]
*2gram261:%x[-1,0]/%x[0,0]
*2gram262:%x[-1,1]/%x[0,1]
*2gram263:%x[-1,2]/%x[0,2]
*2gram264:%x[-1,3]/%x[0,3]
*2gram265:%x[-1,4]/%x[0,4]
*2gram266:%x[-1,5]/%x[0,5]
*2gram267:%x[-1,6]/%x[0,6]

```

```

*2gram268:%x[-1,7]/%x[0,7]
*2gram269:%x[-1,8]/%x[0,8]
*2gram270:%x[-1,9]/%x[0,9]
*2gram271:%x[-1,10]/%x[0,10]
*2gram272:%x[-1,11]/%x[0,11]
*2gram273:%x[-1,12]/%x[0,12]
*2gram274:%x[0,0]/%x[1,0]
*2gram275:%x[0,1]/%x[1,1]
*2gram276:%x[0,2]/%x[1,2]
*2gram277:%x[0,3]/%x[1,3]
*2gram278:%x[0,4]/%x[1,4]
*2gram279:%x[0,5]/%x[1,5]
*2gram280:%x[0,6]/%x[1,6]
*2gram281:%x[0,7]/%x[1,7]
*2gram282:%x[0,8]/%x[1,8]
*2gram283:%x[0,9]/%x[1,9]
*2gram284:%x[0,10]/%x[1,10]
*2gram285:%x[0,11]/%x[1,11]
*2gram286:%x[0,12]/%x[1,12]
*2gram287:%x[1,0]/%x[2,0]
*2gram288:%x[1,1]/%x[2,1]
*2gram289:%x[1,2]/%x[2,2]
*2gram290:%x[1,3]/%x[2,3]
*2gram291:%x[1,4]/%x[2,4]
*2gram292:%x[1,5]/%x[2,5]
*2gram293:%x[1,6]/%x[2,6]
*2gram294:%x[1,7]/%x[2,7]
*2gram295:%x[1,8]/%x[2,8]
*2gram296:%x[1,9]/%x[2,9]
*2gram297:%x[1,10]/%x[2,10]
*2gram298:%x[1,11]/%x[2,11]
*2gram299:%x[1,12]/%x[2,12]

```



## Annexe D

# Propriétés prosodiques associées aux personnages

Cette annexe fournit le détail des mesures moyennes utilisées pour décrire les propriétés prosodiques associées aux personnages, et analysés dans la section [5.5](#). La signification des descripteurs prosodiques utilisés est fournie dans la table [5.2](#).



TABLE D.1: Propriétés prosodiques associées aux personnages

TALE	ID	AGE	GENDER	KIND	VALENCE	PTD	DIIS	IM	NS	NVow	DHIS	HM	TES	PT	TP	PVR
420	Femme	ADULT	F	HUMAN	NEUTRAL	21,0	3,3	71,4	9,0	63	2,8	92,9	10,8	13,3	6,8	1,6
174	Homme Sourd	ADULT	M	HUMAN	NEUTRAL	15,8	3,4	70,4	8,5	221	3,1	86,6	15,8	25,5	6,2	0,0
420	Renard	ADULT	M	FOX	NEUTRAL	18,8	3,3	65,3	7,4	230	4,0	86,6	21,3	16,7	5,7	0,4
21	Héros Fille	YOUNG	F	HUMAN	GOOD	15,2	3,1	66,3	6,8	34	2,2	86,8	9,0	15,7	6,0	2,9
514	Cordon	NA	NA	OBJECT	NEUTRAL	0,0	3,8	70,8	3,5	7	3,5	92,1	11,1	0,0	6,7	0,0
403	Maria	YOUNG	F	HUMAN	GOOD	15,2	2,8	67,2	9,5	46	3,1	89,0	11,1	4,8	6,2	0,0
514	Eau	NA	NA	ELEMENTAL	NEUTRAL	50,0	3,2	70,6	5,0	10	4,9	93,3	17,9	0,0	5,9	0,0
900	Hirondelle	ADULT	NA	BIRD	GOOD	13,7	2,2	71,7	9,0	95	3,1	88,5	14,1	15,7	6,4	0,0
902	Agneau	KID	M	LAMB	GOOD	17,1	2,6	72,5	5,9	41	2,6	92,2	10,0	2,4	6,9	0,0
900	Foule	ADULT	NA	NA	BAD	28,6	3,3	73,6	5,0	35	5,4	90,8	18,9	17,0	5,9	0,0
348	Naëlle	KID	F	HUMAN	GOOD	5,8	2,4	65,2	7,2	122	2,3	88,2	8,2	30,7	6,1	1,6
900	PPMS	KID	M	BAT	GOOD	11,4	2,3	72,5	6,6	45	2,8	91,4	11,8	11,8	6,6	2,2
347	Roi	ADULT	M	HUMAN	GOOD	26,0	4,7	70,1	4,1	77	4,0	92,1	21,8	14,1	3,8	5,2
514	Taureau	ADULT	M	BULL	NEUTRAL	20,0	4,4	71,2	5,5	5	3,1	84,3	5,4	0,0	3,1	0,0
212	PCR	KID	F	HUMAN	GOOD	7,2	2,3	64,0	8,3	141	1,8	90,7	6,7	19,2	5,4	2,1
594	Monstre	KID	M	MONSTER	GOOD	7,0	2,6	67,2	6,8	115	2,0	90,2	8,8	24,4	5,4	0,0
514	Bâton	NA	NA	OBJECT	NEUTRAL	21,4	3,2	72,2	9,0	14	3,2	89,0	16,9	0,0	7,6	0,0
212	Loup	ADULT	M	WOLF	BAD	12,1	2,9	62,1	7,0	147	2,7	82,3	19,7	7,4	4,2	15,6
352	Poule	ADULT	F	BIRD	GOOD	14,2	2,6	68,4	7,5	192	3,1	90,1	17,9	9,7	6,0	1,0
174	Femme Sourde	ADULT	F	HUMAN	NEUTRAL	18,5	2,8	71,7	8,1	130	3,4	88,7	15,7	16,9	6,3	0,0
420	Ours	ADULT	M	BEAR	BAD	16,7	5,2	67,7	6,2	50	4,4	85,0	10,7	9,5	5,5	4,0
352	Chien	ADULT	M	DOG	NEUTRAL	13,3	2,7	68,5	10,0	167	2,6	88,5	12,5	8,8	6,1	0,6
348	Louveau	KID	M	WOLF	GOOD	0,0	2,3	69,8	7,5	15	2,8	91,4	7,7	7,6	5,7	0,0
21	Mère1	ADULT	F	HUMAN	BAD	18,8	3,3	71,5	7,2	129	3,3	88,5	16,3	18,9	6,1	0,8
514	Feu	NA	NA	ELEMENTAL	NEUTRAL	22,2	2,4	70,3	4,5	9	3,2	92,3	14,5	0,0	6,8	0,0
348	Chef Loup	ADULT	M	WOLF	BAD	18,1	3,6	71,1	7,4	74	2,4	85,8	11,6	14,8	6,3	2,7
902	Chevreau	KID	M	GOAT	GOOD	22,5	3,5	72,2	4,4	40	4,3	91,6	15,4	7,3	6,1	0,0
514	Couteau	NA	NA	OBJECT	NEUTRAL	28,6	2,5	75,7	3,5	7	3,6	93,7	14,7	0,0	4,6	0,0
514	MaitresseDeMaison	OLD	F	HUMAN	BAD	10,5	3,2	70,0	14,4	499	2,0	88,9	11,8	7,2	6,7	1,0
347	Conseiller	ADULT	M	HUMAN	GOOD	17,9	2,3	68,8	8,0	56	4,8	86,3	17,6	12,7	5,1	0,0
347	Ministre	ADULT	M	HUMAN	GOOD	38,5	3,5	72,0	4,3	13	3,0	93,3	7,3	16,2	4,5	0,0
348	Papa Oiseau	ADULT	M	BIRD	GOOD	7,1	3,8	70,3	9,3	28	2,2	90,7	17,6	4,3	8,0	0,0

TABLE D.1: Propriétés prosodiques associées aux personnages

TALE	ID	AGE	GENDER	KIND	VALENCE	PTD	DIIS	IM	NS	NVow	DHIS	HM	TES	PT	TP	PVR
902	Corbeau	ADULT	M	BIRD	GOOD	14,3	3,4	69,9	4,0	28	4,8	91,9	18,3	14,8	5,1	0,0
352	Fermier	ADULT	M	HUMAN	NEUTRAL	14,5	3,3	71,0	9,7	63	2,0	90,5	10,2	7,9	6,5	1,6
403	Fée1	ADULT	F	FAIRY	GOOD	16,7	2,7	61,9	6,5	85	2,9	82,4	14,6	12,7	5,9	1,2
347	Suzon	ADULT	F	HUMAN	GOOD	8,7	3,0	73,5	11,5	23	2,3	91,5	9,2	2,5	5,7	0,0
900	Maman Souris	ADULT	F	MOUSE	GOOD	18,1	2,3	71,3	7,0	199	3,0	88,6	13,4	17,3	5,9	0,0
403	Mère2	ADULT	F	HUMAN	NEUTRAL	8,3	2,8	65,1	8,1	73	3,4	85,0	12,0	10,6	7,4	1,4
902	Fermier	ADULT	M	HUMAN	BAD	28,1	4,2	74,7	4,1	33	6,0	91,7	17,8	9,8	4,9	3,0
902	Grand Mère	OLD	F	BULL	GOOD	26,3	2,7	71,6	5,4	19	5,8	87,3	21,9	3,2	4,9	0,0
514	Brirouch	YOUNG	M	HUMAN	NEUTRAL	40,0	4,1	69,0	2,5	5	5,0	90,4	12,8	0,0	5,1	0,0
594	Maxence	KID	M	HUMAN	GOOD	10,9	2,6	70,6	6,6	159	2,6	90,1	11,5	19,3	5,8	1,9
All	N0	N0	N0	N0	N0	12,7	2,9	66,0	7,4	8059	3,4	84,2	16,1	24,0	6,2	2,5
902	Fermiere	ADULT	F	HUMAN	GOOD	17,2	3,3	73,0	6,9	58	3,4	89,7	17,4	18,2	6,6	0,0
514	Forgeron	ADULT	M	HUMAN	NEUTRAL	14,3	3,0	70,6	3,0	9	3,7	84,9	11,1	0,0	5,5	22,2
348	PapaOurs	ADULT	M	BEAR	GOOD	25,0	3,2	70,7	8,0	24	2,7	84,8	13,3	11,7	5,8	0,0
900	Engoulevent	ADULT	NA	BIRD	GOOD	11,8	2,8	72,3	8,0	68	3,0	87,6	14,6	14,6	6,4	0,0
902	Veau	KID	M	BULL	GOOD	14,4	2,6	72,0	6,1	146	2,8	92,4	10,2	8,0	6,5	0,0
347	Medecin	ADULT	M	HUMAN	GOOD	20,0	1,9	71,5	10,0	10	1,9	91,2	6,2	0,0	4,9	0,0
352	Chat	ADULT	M	CAT	NEUTRAL	15,3	2,4	69,7	9,0	151	2,4	89,5	13,5	12,5	6,3	0,7
21	Soeur	YOUNG	F	HUMAN	BAD	7,7	3,1	70,2	7,2	65	2,4	89,2	11,8	18,7	6,8	0,0
902	Maman	ADULT	F	BULL	GOOD	40,0	2,6	74,1	6,3	15	4,2	91,3	14,7	6,2	4,2	0,0
174	Chef	ADULT	M	HUMAN	NEUTRAL	12,8	3,9	69,3	7,8	78	4,5	86,6	18,0	24,4	6,5	0,0
348	Grand Mère2	OLD	F	HUMAN	GOOD	17,1	3,3	67,7	5,8	70	3,9	86,1	16,6	22,7	6,6	0,0
420	Erik	ADULT	M	HUMAN	NEUTRAL	11,3	2,7	69,5	7,5	159	2,3	88,7	14,2	12,4	6,2	0,0
514	Chat	ADULT	M	CAT	NEUTRAL	25,0	5,9	64,4	4,0	4	6,2	88,0	19,4	0,0	4,7	0,0
900	PPM1	KID	M	BAT	GOOD	16,7	2,1	69,9	3,0	6	6,5	94,5	7,6	0,0	4,3	0,0
21	Fée2	ADULT	F	FAIRY	GOOD	7,2	2,5	66,3	6,1	98	2,4	84,8	13,6	21,5	6,5	1,0
900	PPM3	KID	M	BAT	GOOD	28,6	3,5	68,9	3,5	7	3,8	92,4	7,7	0,0	4,0	0,0
900	PPM2	KID	M	BAT	GOOD	11,1	3,1	67,2	5,0	10	4,9	94,7	19,8	0,0	4,6	10,0
212	Grand Mère3	OLD	F	HUMAN	GOOD	35,7	3,7	68,7	4,7	14	5,7	91,3	24,6	5,8	4,5	0,0
514	Rat	ADULT	M	RAT	NEUTRAL	0,0	4,7	72,8	3,0	6	3,9	92,3	12,5	0,0	5,4	0,0
212	Mère4	ADULT	F	HUMAN	GOOD	9,7	2,6	68,2	10,7	32	3,0	87,1	15,7	15,7	5,9	3,1
352	Paon	ADULT	M	BIRD	BAD	15,7	3,0	70,9	8,3	108	3,8	91,1	20,0	17,0	5,8	0,0



## Annexe E

# Textes sélectionnés pour l'évaluation du système de synthèse

### E.1 Erik , le paysan rusé

Erik , le paysan rusé

un jour , un paysan était allé dans la forêt avec son cheval et son traîneau pour faire la provision de bois .

il n' avait pas terminé son ouvrage qu' un énorme ours s' approcha de lui et lui dit :

- donne -moi ton cheval sinon cet été , tu pourras faire attention à tes moutons .

- oh l' Ours !

répondit le paysan , il ne me reste plus de bois à la maison , pas une seule petite bûche .

je t' en prie , laisse -moi ramener cette charge de rondins jusque chez moi .

tu ne voudrais pas que je meure de froid ?

demain , je te promets de te ramener le cheval .

### E.2 Crétin de paon

- je ne m' appelle pas Piwi , répondit orgueilleusement le paon en faisant la

roue et le cou allongé , en criant d' une voix aigre : mon nom est " sa - queue

- est - magnifique - comme - le - soleil - quand - il - brille - sur - les -  
montagnes - dans - les - brouillards - du - matin . "

est -ce compris ?

désormais ne m' appelle plus autrement , sans quoi je t' arracherai les yeux .

- je ne t' appellerai jamais autrement , promis la poule , qui redoutait le bec acéré de l' autre , jamais , jamais .

### E.3 Naëlle va chercher du bois dans la forêt

le chef regarde d' où vient la voix et se moque du Papa oiseau qui vole au-dessus de lui avec Maman oiseau et Petit oiseau .

- " ce n' est pas toi qui va me faire peur ! "

c' est à ce moment qu' il entend une grosse voix qui lui demande :

- " et moi , je te fais peur ? "

il se retourne et voit Papa ours avec Maman ours et Ourson qui sont vers lui .

- " cette petite fille a sauvé mon fils qui avait la patte prise dans un piège .  
"

" elle a aussi donné à manger à mon fils qui a pu nous retrouver " dit Papa oiseau .

### E.4 Le roi Glagla

le roi Glagla éternuait si fort que les murs du château tremblaient .

des jours durant , il éternua .

puis un matin le roi ne voulut pas quitter son lit , pas même pour aller faire pipi .

il était tout blanc et claquait des dents :

- je , je ... je suis ge ... gelé .

son ministre s' affola :

- au secours , à moi !

notre bon roi se meurt de froid !

## E.5 L'histoire des trois sourds

la Leçon est :

il vaut mieux ne pas se dépêcher de donner une réponse .

on conseille quelque part en Afrique , d' avoir le cou aussi long que celui du chameau , afin que la parole avant de jaillir puisse prendre tout son temps .

## E.6 Maxence et le monstre sous le lit

le monstre rose hésita un peu , puis grimpa avec appréhension sur le lit , en s' aidant de sa longue queue .

- c' est bien vrai , il n' y a aucun monstre sur ce lit !

mais je ne suis quand même pas bien rassuré ...

dis moi petit garçon , puis -je dormir avec toi cette nuit ?

j' aurais moins peur , si tu es là pour me protéger .



## Annexe F

Stylisation du conte “Le petit  
chaperon rouge” contenu dans le  
corpus de parole GV-LEx



